PUBLICATIONS OF THE SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION

XIX

INTERNATIONAL EXAMINATION INQUIRY

SELECTION FOR SECONDARY EDUCATION

SELECTION FOR SECONDARY EDUCATION

BY

WILLIAM McCLELLAND

UNIVERSITY OF LONDON PRESS LTD.
WARWICK SQUARE, LONDON, E.C.4
1949

Reprinted 1945, 1949

Printed in Great Britain by
NEILL & CO., LTD., EDINBURGH.

INTERNATIONAL EXAMINATION INQUIRY COMMITTEE

Convener

JAMES DREVER, M.A., B.Sc., D.Phil., Professor of Psychology, Edinburgh University

Committee

WILLIAM BOYD, M.A., B.Sc., D.Phil., D.Litt.,	Reader in Education, Glasgow University
D. KENNEDY-FRASER, M.A., B.Sc.,	Jordanhill Training College, Glasgow
*THOMAS HENDERSON, B.Sc., F.R.S.E.,	General Secretary, The Educational Institute of Scotland
W. A. F. HEPBURN, O.B.E., M.C., M.A., B.Ed., LL.D., F.R.S.E.,	Director of Education for Ayrshire
JOHN MACKIE, M.A., D.Sc., F.R.S.E.,	Rector, Leith Academy
WILLIAM MCCLELLAND, M.A., B.Sc., B.Ed.,	Executive Officer of the National Committee for the Training of Teachers (formerly Professor of Education in the University of St Andrews, and Director of Studies, Dundee Training College)
R. R. RUSK, M.A., B.A., Ph.D.,	Director to the Scottish Council for Research in Education
J. C. SMITH, C.B.E., M.A., Litt.D., LL.D.,	Sometime H.M. Senior Chief Inspector of Schools for Scotland
GODFREY H. THOMSON, Ph.D., D.Sc., D.C.L.,	Professor of Education, Edinburgh University
NORMAN T. WALKER, M.A., Ed.B., Ph.D.,	Lecturer in Education, Aberdeen University
JOHN WISHART, M.A.,	General Secretary, The Educational Institute of Scotland

* Died 9th March 1941

LIST OF THE MAIN VOLUNTARY ASSISTANTS IN THE INQUIRY

I. *Chief Assistants*

Miss Margaret Young, M.A., B.Ed., Lecturer in Experimental Education, Dundee Training College

Douglas M. McIntosh, M.A., B.Sc., B.Ed., Ph.D., Assistant Director of Education for the County of Fife

II. *Group Leaders; Instructors; Assistants in more Advanced Statistical Methods*

James C. Kidd, M.A., B.A., Boroughmuir Secondary School, Edinburgh

Miss Muriel Mitchell, M.A., Dundee Training College

Alex. S. Robertson, M.A., Harris Academy, Dundee

John T. Bain, M.A., B.Sc.
Frank Esplin, M.A.
Miss Jane Fairley, B.Sc.
Miss May Herald, M.A.
Alfred McLean, M.A.
William M. Williamson, B.Sc.

III. *Students who conducted Special Investigations* [1]

John T. Bain, M.A., B.Sc.
George A. Crabb, B.Sc.
David C. Johnstone, M.A.
James C. Kidd, M.A., B.A.
Alex. S. Robertson, M.A.
Miss Jeanie I. W. Robertson, M.A.
Robert Robertson, M.A.
Miss Hilda C. Rodgers, M.A.
James P. Shepherd, M.A.
Hugh Stewart, B.Sc.
Robert Wilkie, B.Sc.

[1] The list includes only those whose results have been used in the Report.

IV. *Teachers who assisted in Correction of Tests, Tabulation, Calculations, etc.*

John W. Baxter, M.A.
George L. Boyd, M.A., B.Ed.
Miss Katharine R. Brebner
Miss Margaret Burnett, M.A.
Miss Isobel V. Calder
Miss Marion W. Coupar, M.A.
Miss Annie N. Dewar, M.A.
Miss Hilda I. Drummond, M.A.
Miss Ellinor J. Dye
Miss Felicia K. Fernie, M.A.
Miss Isabella M. Glen, M.A.
Miss Annie M. Gow
Miss Muriel M. H. Hunt, B.Sc.
Robert J. Jack
John J. Kitchin
William C. Laurie
Miss Barbara L. Leah
Miss Annie M. Leslie
Miss Elizabeth McCaskill
Miss Katherine M. A. MacGregor
Miss Annie McKay
Miss Jane A. Mackay, M.A.
Miss Mary G. Mackenzie, B.Sc.
Alexander Matheson
Miss Adelaide Neilson
Miss Helen C. Neilson
Miss Frances E. Reynolds
Miss Wilhelmina Simpson
Miss Evlyn I. Soutar, B.Sc.
Alexander F. Stark, M.A.
Henry Sturrock
Miss Elizabeth M. Sutherland
Miss Floretta Whitely, B.Sc.
Miss Joyce Wilcox, M.A.
Miss Isabella Winter, B.Sc.
Miss Ethel Wishart
Miss Jean M. Yule, M.A.

FOREWORD

The present publication represents the final Report of the Scottish Committee, or Delegation, in the International Examination Inquiry initiated by the Carnegie Corporation of America, the Carnegie Foundation and the International Institute of Teachers College, Columbia University, and financed jointly by the Carnegie Corporation and the Carnegie Foundation. It also represents the completion of the tasks undertaken in 1931 at the Eastbourne Conference by the Scottish Delegation. This, the fourth of our major investigations, had not defined itself very clearly until just before the second Conference, held at Folkestone in 1935, where the plans were submitted for criticism and for approval. The third Conference, at Dinard in 1938, saw the investigation so far advanced that Professor McClelland was able to give a detailed account of what had been completed and what still remained to be done. Now, four years later, he is able to present to the Committee and to the public the full results. It will, I think, be generally agreed that this has been the most important and the most onerous of all the investigations sponsored by the Committee; it consequently seems incumbent on me as Convener to express our indebtedness to Professor McClelland, who has throughout borne on his own shoulders practically the whole burden. The measure of that indebtedness can be assessed by the reader of this work.

At the Eastbourne Conference the Chairman, Dr Paul Monroe, stressed the sociological implications of examinations, and pointed out that these implications constitute the most general aspect of the whole series of investigations undertaken by the various national delegations. The enormous sociological significance of the present investigation has been indicated by Professor McClelland, and any elaboration is superfluous. It is worth noting, however, that the Mental Survey investigations of 1932 and 1935–37 might be said to represent the first stage in a great educational and sociological study, of which the present investigation represents the second stage, and that a third stage remains to be undertaken—a mental and educational survey of the population of our secondary and technical schools, our universities and central institutions, as

well as our adult education classes. This third stage has already been envisaged by the Scottish Council for Research in Education, but financial considerations preclude for the time being the possibility of undertaking an investigation so complex, so difficult and so costly.

Although the International Examination Inquiry as initiated and financed by the International Institute of Teachers College has now terminated, at least so far as the Scottish Delegation is concerned, we are still faced with this great task, and with the study of other problems concerned with examinations. The hope may be expressed that, when the present difficult days are past, it will again be possible to secure international collaboration in this work. At the closing session of the Dinard Conference in 1938, on the motion of the Convener of the Scottish Delegation, it was unanimously agreed that an international committee for the study of examinations be appointed, and that it consist in the first instance of the chairmen of the existing committees. Subsequently Professor Godfrey H. Thomson consented to act as secretary to such a committee. Unfortunately the unsettled state of the Continent, and then the War, made it impossible to carry the organisation of the committee beyond the skeleton stage. When peace returns, it may not be too much to expect that this committee will begin to function, and continue the work so successfully prosecuted for seven years under the ægis of the International Institute, and the inspiring leadership of Dr Paul Monroe.

It only remains to express once more in the name of the Scottish International Examinations Inquiry Committee and the Scottish Council for Research in Education their sense of the deep debt they owe—and Scottish education owes—to the International Institute of Teachers College for the continued encouragement they have received, and for the very substantial financial assistance without which the investigation could not have been undertaken and published.

JAMES DREVER,
Convener.

PREFACE

This work is a scientific study of certain problems of selection and guidance that arise at the first big sifting of the material in an educational system; and, as such, it is not directly concerned with many of the deeper issues of educational and examination policy. We set out to find answers to a number of definite questions in which administrators are interested, but we offer no opinion as to whether, in a sound educational system, the questions would arise in these particular forms, or would, indeed, arise at all.

Our thanks are due to the Convener and Members of the Dundee Education Committee not merely for their ready and generous response to our application for facilities to conduct the experiment, but also for the encouraging interest which they have taken in its progress. To their Director of Education, Mr John R. Cameron, M.A., we have been indebted for the preparation of the examination papers and for constant guidance and help in all our difficulties. Dr A. E. Kidd, M.B., C.M., D.P.H., former Chief Medical Officer for Schools, and his staff undertook the heavy task of furnishing us with health gradings and medical reports for the complete group of over 3,000 pupils; and Dr Kidd also prepared the classification of schools on the basis of social class. We have also to record our appreciation of the co-operation of the head teachers and staffs of the City schools, to whom we are deeply grateful for the friendly and helpful way in which they met our heavy demands upon their time and patience.

Lengthy as the Report is, it is only to those with actual experience of similar research that it will convey a full realisation of the labour involved. The determination of a single correlation coefficient, when the numbers are large, is a long and laborious operation; and, to ensure accuracy, all our calculations were done twice, by different teams. With this duplication the number of direct correlation calculations in the main Inquiry was over 3,000: the testing programme involved the correction of over 20,000 scripts.

To embark upon such an undertaking without a large full-time staff implied a belief, which events have proved to be well-founded, that we could safely count upon the help of our students and of

our colleagues in the City schools. The Inquiry was, in fact, an adventure in large scale co-operative research carried out by voluntary part-time assistants; and a brief description of the organisation may be of interest to other investigators.

My two partners in the general conduct of the Inquiry were Miss Margaret Young, M.A., B.Ed., Lecturer in Experimental Education in Dundee Training College, and Douglas M. McIntosh, M.A., B.Sc., B.Ed., Ph.D., Assistant Director of Education for the County of Fife. Miss Young's special responsibilities included charge of the testing, preparation of circulars of instructions, training teams of testers and correctors, etc. She is the writer of Chapter I, in which she gives an account of these matters. Dr McIntosh, who has written Chapter II, undertook such tasks as the standardisation of the tests, organisation of the correlation calculations, training teams of computers, and so on. Yet neither a list of special duties nor the occasional references in the Report can give an adequate appreciation of what the Inquiry owes to these two principal assistants. Each envisaged the research as a whole and helped with practically every aspect of it.

So also did Miss Muriel Mitchell, M.A., who, before her appointment to the administrative staff of the College, had charge of the rapidly mounting mass of documents in the room set aside for the Inquiry. Since 1936 she has given voluntary help of the utmost value and has been much in request as a consultant to whom student assistants went singly or in groups for advice in their statistical difficulties.

The correction of the tests and the correlation calculations were carried out by a veritable army of helpers, over 600 in all, in whose ranks were members of the College staff, head teachers and teachers of the City schools, and students of the Training College. Most of them had to be trained for the work, and this training was given in the first place by Miss Young and Dr McIntosh. Later, certain of the more skilled workers acted as group leaders and trained new teams. Two of these group leaders, Alex. S. Robertson, M.A., and James C. Kidd, M.A., B.A., gave particularly valuable help with the more difficult calculations and many special problems.

The teams of correctors and computers fell roughly into two types, the 'permanent' and the 'occasional.' Into the former came a most faithful group who worked from 6 to 9 p.m. on one or two evenings a week for several sessions. In the latter we had groups of students who worked in forenoons during vacations;

others who worked in free periods during the College session; others who worked individually at home, coming up to the evening meetings periodically to discuss difficulties.

Finally, about fifty graduate students undertook studies of special aspects. Each year, after a talk on the aims and progress of the Inquiry, a list of problems was submitted from which interested students could make their choice. There was no lack of volunteers for these investigations, many of the results of which will be found in the Report.

It may not be out of place to say a word as to the part played by the Inquiry in the life of the Training College. The students' help was invaluable, and our gratitude to them is not diminished by the belief that the benefits were mutual. For many of them simple statistics forms part of the course of training, and the calculations which they carried out in connection with the Inquiry merely replaced class-room exercises which would have had less reality and interest. Moreover, while all types of student participated, they were not asked to work in a blind mechanical manner. Before setting to work they understood the plan of the research and the way in which their particular calculation fitted into the whole. Whenever possible we let them know the results which their group had obtained, and discussed their significance. In short, the Inquiry was a central research interest for the whole College for six years. Its general headquarters, the Inquiry room, was a meeting-place for students from all sections who were co-operating in the work. It had a real research atmosphere, and no one could associate with the students in it without a feeling that the project was an enriching influence in the life of the College.

Of the many colleagues who helped us we are specially indebted to Mr B. Babington Smith, M.A., of the Psychology Department of the University of St Andrews, with whom we had many fruitful discussions on statistical problems, and to the following members of the staff of the Dundee Training College: Mr Andrew Nairn, Principal Lecturer in Art, who prepared the diagrams; Mr Robert B. Martin, Chief Clerk, who acted as Treasurer; Miss Dorothy Mess, and other members of the administrative staff whose help included the typing and duplicating of circulars, etc.; and Mr George Caithness, janitor, who helped us in ways too numerous to specify.

Throughout the Inquiry, from the original formulation of the plan to the final preparation of the Report, we had the benefit of

the expert advice of the Research Council's Director, Dr R. R. Rusk. The extent of our indebtedness to him will be best appreciated by those who have carried out researches for the Council.

Finally, a reference may be made to the Report itself. It is primarily a research report; and the ways of educational research have become very technical: yet we have tried to present it in such a way that the results will not be rendered inaccessible to the practical administrator through being expressed in the unintelligible mathematical jargon that would be more pleasing to the expert. We have kept the mathematics as far in the background as possible, and we believe that the lay reader, whose statistical attainments are limited to a knowledge of the meaning of the terms 'standard deviation' and 'correlation coefficient,' will be able to follow the text. He should simply ignore the footnotes. These are for those who are interested in the techniques; but to avoid possible disappointment, it should be mentioned that most of our methods and formulæ had to be devised *ad hoc*, and that the references often give merely the starting-point from which it will be possible to retrace our mathematical steps.

We have tried to throw a little light on some of the dark places of the selection problem: but by the time the reader has finished the Report he will realise that the torch has been a heavy one to carry—and we now willingly hand it on to others.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

PLAN AND ORGANISATION OF THE INQUIRY

The Qualifying Stage in Scottish Education

What is known as the Qualifying stage in Scottish education is that at which a pupil is transferred from the uniform primary course to differentiated post-primary courses. The normal age of transfer is 12 years.

Until December 1921 all pupils who had completed a primary course had to undergo a Qualifying examination, conducted by H.M. Inspectors, the function of which was to ascertain whether the pupil had satisfactorily completed the primary stage and was fit to enter a post-primary course. When this examination was abolished, the Scottish Education Department indicated, in *Circular No. 44*, that the Qualifying stage 'must continue to mark something of an epoch in the school career of every pupil'; and Education Authorities were advised to substitute for the Department's Examination 'some arrangement under which those who have taught and those who are to teach particular individuals shall combine in an endeavour to estimate the potentialities of the material to be handled.'

Since 1922 there has been a considerable amount of experiment with different methods of complying with this injunction. Each Education Committee has evolved its own scheme of promotion at the Qualifying stage, and these schemes are modified from time to time in the light of experience gained. They range from simple systems where only one measure of attainment is used, to elaborate systems involving the use of standardised mental and scholastic tests, ordinary examinations and teachers' estimates.

The uses which are made of the results of the examinations at the Qualifying stage vary from Area to Area.

The Object of the Inquiry

The aim of this Inquiry is to study the problems of selection and guidance that arise at the Qualifying stage, and to find the system of assessment of ability and attainment which will give the

most reliable forecast of the type of post-primary course for which each pupil is fitted. In particular the object is to find answers to the following questions:

What are the relative predictive values of (*a*) ordinary examinations, (*b*) teachers' estimates, (*c*) group intelligence tests and (*d*) standardised scholastic tests?

What are the best combinations of the above measures in selecting pupils for secondary courses?

The General Plan of the Inquiry

Satisfactory answers to such questions can be obtained only by means of a large-scale follow-up investigation. We therefore first tested the complete Qualifying group in a Scottish City by all the methods of assessment at present in use, obtaining additional data on special rating cards. The progress of these pupils in the various post-primary courses was then followed for three years.

We proposed to continue the follow-up of the senior secondary pupils for five years, but this had to be abandoned because of war conditions.

THE QUALIFYING GROUP

The Qualifying group consisted of all pupils presented for promotion to post-primary courses in Session 1935–1936. The total number was 3,229, of whom 1,130 were presented in December 1935 and 2,099 in June 1936. In view of the two dates of presentation, duplicate tests and examinations had to be prepared. The age distributions of the two groups were as follows:

Age	December 1935		June 1936	
	N	Per cent.	N	Per cent.
14–15	3	·3	9	·4
13–14	46	4·1	126	6·0
12–13	578	51·1	1,018	48·5
11–12	501	44·3	940	44·8
10–11	2	·2	6	·3
	1,130		2,099	

THE AGE-GROUPS

The age-groups were used for the standardisation of the intelligence and scholastic tests. They consisted of all pupils in

the City of ages from 11 years 6 months to 12 years 5 months on the dates of the two examinations. The distributions of these age-groups among the classes were as follows:

Class	December 1935 N	December 1935 Per cent.	June 1936 N	June 1936 Per cent.
Primary				
III	48	1·6	45	1·5
IV	249	8·1	348	11·9
V (Qualifying class) .	1,860	60·3	1,940	66·7
Secondary				
First year . . .	906	29·4	576	19·8
Second year . .	19	·6	2	·1
	3,082		2,911	

THE QUALIFYING DATA

Standardised Tests

The following Moray House Tests prepared by Professor Godfrey H. Thomson were used: [1]

Subject	Test: December 1935	Test: June 1936	Duration in minutes: Practice	Duration in minutes: Test	Duration in minutes: Total
Intelligence .	MHT 20	MHT 21	10	45	55
English . .	MHE 7	MHE 8	..	40	40
Arithmetic .	MHA 7	MHA 8	..	30	30

Qualifying Examinations

Subject	Description	Duration in minutes
English . .	Dictation and Spelling .	30 approximately
	Composition	35
	General English	
	(*a*) Interpretation of passage	
	(*b*) Simple questions on English usage	60
Arithmetic . .	Mental	10
	Mechanical and problems . . .	60

[1] Particulars of the Moray House Tests may be found in *What are Moray House Tests?* London: University of London Press, Ltd., 1940.

The papers, which were based upon those set in recent years to Qualifying pupils in various Scottish counties, were prepared by the Director of Education for the City. Copies of the December papers are given in Appendix A.

Teachers' Estimates of Attainment in the Primary School

The head masters and teachers were in the habit of estimating the attainment of each pupil at the Qualifying stage, and these estimates were obtained from the schedules sent to the Director of Education. Marks were on a percentage basis in English and Arithmetic, while a qualitative grading was given in practical subjects such as Handwork and Drawing.

Bursary Examinations

The City Education Committee allowed the results of the bursary examinations in English and Arithmetic set to applicants for bursaries for senior secondary courses to be used.

Rating Cards

A rating card was devised for the purpose of securing qualitative estimates of each child's health, ability, attainment and personality. The primary head masters were responsible for completing this card, guided by a carefully prepared memorandum of instructions. The grading for health was supplied by the Chief Medical Officer for Schools. This rating card is shown in Appendix B.

ADMINISTRATION OF THE TESTS AND EXAMINATIONS

Compilation of a Card Index

Particulars regarding each pupil were supplied by head masters on special cards. These cards provided an estimate of numbers in each school and particulars for the tabulation sheets which had to be prepared before the tests and examinations were corrected.

Arrangements in Schools

Head masters co-operated in securing the best possible testing conditions. Pupils were generally tested in groups of 25 to 30, seated at single desks. Circulars were prepared to ensure uniformity, and in deciding matters relating to internal school arrangements the assistance of the Director of Education was invaluable. The following circulars to head masters were issued:

Circular No. I. Mental and Scholastic Tests. This gave instructions regarding time-table, arrangement of pupils in

groups, names of examiners and supervisors, material required, return of material.

Circular No. II. Mental and Scholastic Tests. This brief memorandum confirmed the times stated in Circular No. I and gave full details of the duration of the tests.

Circulars of a similar type were issued in connection with the examinations.

Time-table of Tests and Examinations

The intelligence and scholastic tests were given on three consecutive days. The examinations were given on two consecutive days about a fortnight after the last test.

Distribution and Collection of Material

Great care was exercised in the distribution and collection of standardised test and examination papers. Sealed packets were delivered to each head master and handed by him to an accredited supervisor before each test period. The standardised tests had not previously been used in the City. All used and unused tests, all examination question papers and examination scripts, were returned to the Dundee Training College at the close of each day's testing.

Teams of Testers and Supervisors

The standardised tests were administered by teams of teachers and students trained in this technique. The same teams conducted the Qualifying examinations under carefully controlled conditions. Meetings were arranged with teachers and students at which details of procedure were discussed, and an accredited supervisor was appointed to each school to make contact with the head master. The following circulars were distributed to the testers and supervisors:

Memorandum for Testers No. I. Mental and Scholastic Tests. Details were given of dates and duration of test periods, and instructions as to test procedure were carefully stated.

Memorandum for Testers No. II. Qualifying Examinations. Details were given as to distribution of examination material, heading of answer sheets, and administration of examinations.

Memoranda for Supervisors of Tests and Examinations. These memoranda gave instructions for the distribution and collection of test material, examination material and absentee slips.

The care exercised in the conduct of the ordinary type of examinations was occasioned by the belief that such examinations may be at a disadvantage relative to standardised tests if they are given under ordinary school conditions while the tests are administered under rigorous research conditions.

Correction and Tabulation of Results

(*a*) Standardised tests. The efficient administration of the tests and examinations was due to the loyal team work of head masters and teachers of primary and secondary schools assisted by Training College students. Their continued co-operation made light work of the formidable task of correcting the 10,000 standardised tests. The correction, checking, calculation of quotients and tabulation of results were carried out in the Training College by teams of volunteers under supervision.

(*b*) Qualifying examination scripts. The Qualifying examination scripts were marked by a paid examiner who has for many years corrected the Qualifying examination papers in a Scottish Education Area.

THE FOLLOW-UP

The System of Promotion to Post-Primary Education

At the time of the testing in Session 1935–1936 promotion to post-primary education in the City was based on the estimates of the pupils' attainment given by the primary schools. On this basis the Education Committee decided which pupils should be transferred to special classes for backward children or retained in primary schools for later presentation. Pupils regarded as fit for promotion to three-year or five-year secondary courses were guided by the head teachers of the primary schools as to choice of course, but there was no compulsion.

The only selection for five-year secondary education resulted from the fact that, at this time, fees were charged in all the senior secondary schools. Pupils desiring a full secondary education, who would have been debarred by reason of the expense involved, took a special examination. If they attained a certain standard in this examination, defined from year to year by the Education Committee, they were admitted free to a senior secondary school.

The results of the standardised tests and Qualifying examinations given in this investigation were *not* used to influence promotion.

Tracing the Qualifying Group

The tracing of the 3,229 pupils presented for the Qualifying examination in 1935–1936 proved to be an intricate task. The courses to which they proceeded are shown below:

Course	N	Per cent.
Senior secondary courses	461	14·3
Junior secondary courses		
Boys' technical	870	26·9
Girls' technical	763	23·6
Commercial	379	11·7
Special classes for backward pupils	357	11·1
Retained in primary schools with a view to later presentation	342	10·6
Others: left City, etc.	57	1·8
	3,229	

Record Booklets

The *Dundee Post-Primary School Record Booklets* were used to record the educational history of the pupils concerned in the investigation. The original booklets were sent to the Training College at the end of each examination period and a duplicate record was made of the marks of each pupil.

Final Follow-up Cards

Final follow-up cards were devised for the purpose of securing a general estimate of the pupil's progress at the conclusion of the first three years of the post-primary course, or when he left school. Three types of card were required: for pupils in senior secondary schools; for pupils in junior secondary schools; for pupils in backward classes. Detailed instructions for completing these final follow-up cards were sent to the head masters of the schools concerned. A follow-up card is shown in Appendix C.

CHAPTER II

STANDARDISATION OF THE TESTS

THE results of intelligence and scholastic tests may be used either in the form of scores or of quotients; that is to say, without or with an allowance for differences in the ages of the candidates. As one of our aims was to find whether the score or the quotient should be used in selection, we required a reliable means of transforming scores into quotients, and this involved the standardisation of all the tests.

REASONS FOR NOT USING PREVIOUS STANDARDISATIONS

The Moray House Tests which were used had already been standardised for an English county. At first it was thought that these standardisations might be suitable for the purpose of the Inquiry, but there were several objections to this procedure.

Firstly, there was doubt as to whether the results from an English county with a large rural population would be similar to those from a Scottish industrial city. Again, the conditions under which tests are given in England differ from those under which such an experiment is carried out in Scotland. For example, it is a general rule in England that these tests are given to determine which children shall obtain free places in a secondary school, and this gives the candidates an incentive to do well. Added to this is the fact that the tests have been given for several years in the county, with the result that the children have acquired a certain familiarity with the form of the questions.

These difficulties had already been experienced by the Scottish Council for Research in Education when it undertook a national survey of a complete age-group in 1932.[1]

The comparability of the norms might also be affected by the fact that the groups selected for the standardisation of the tests in England were at a lower age-level than that of the group tested in the Inquiry.

[1] *The Intelligence of Scottish Children.* Publications of the Scottish Council for Research in Education No. V. London: University of London Press, Ltd., 1933. Pp. 72–79.

From a previous survey it was found that the average age of the Qualifying pupils in the City was 12 years. As the ages of 73 per cent. of the pupils lay within the age-group 11 years 6 months to 12 years 5 months it was decided to standardise the tests on this group, and to extend the norms on either side of the age-limits.

The distributions of the children in the age-groups for December and June have already been given on p. 3.

Method of Standardisation

The tests were standardised by a method devised by Professor Godfrey H. Thomson.[1]

Comparison of Inquiry and English Results

The previous standardisations enable us to make some comparisons between Inquiry and English children in regard to intelligence and scholastic attainment. If Scottish readers find the results to be somewhat unexpected, they must remember, in addition to the special explanations already given, that the intelligence level of a country is not uniform.

INTELLIGENCE TESTS

The comparison on the basis of MHT 20 is given in fig. 1, which will be intelligible to those who are not familiar with Professor Thomson's technique of standardisation. The lines may be regarded as giving the average scores of pupils of different ages at various levels of ability.

The first point to be observed is that the groups on which the standardisations are based are at different age-levels, the English group ranging from 10 years 9 months [2] to 11 years 8 months while the Inquiry group range from 11 years 6 months to 12 years 5 months. The most outstanding feature of the diagram is the superiority of the English group at every level of ability. Another feature is the parallelism of the corresponding lines, indicating that the age-allowances at the different levels are in close agreement. When translated into terms of IQ the difference between the two

[1] G. H. Thomson, "The Standardisation of Group Tests and the Scatter of Intelligence Quotients," *British Journal of Educational Psychology*, vol. ii, Part II, June 1932, pp. 92–137.

[2] Only that part of the range which is relevant to the comparison is shown in the figure.

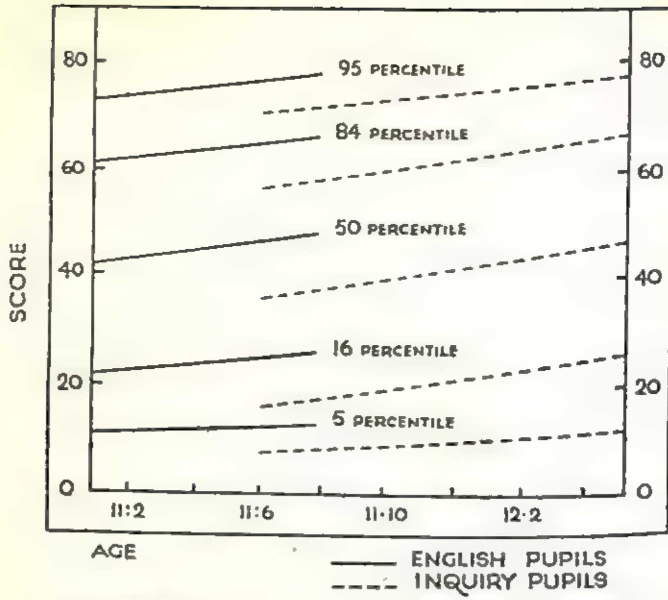

FIG. 1.—Intelligence: Comparison between English and Inquiry pupils (MHT 20).

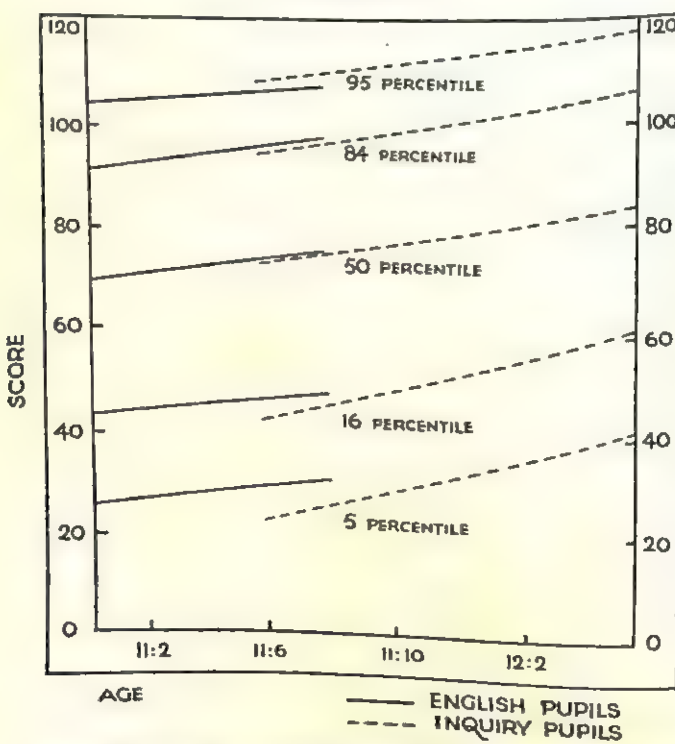

FIG. 2.—English: Comparison between English and Inquiry pupils (MHE 7).

standardisations amounts to something in the region of 7 to 8 points.

At first sight it would appear that the English children are superior in intelligence to those of the Inquiry group, but several important points must be considered before any such conclusion can be drawn. These, as already mentioned, include the familiarity of the English children with the form of questions and the incentive to do well in such tests.

It is important to note that the process of standardising the test on the City age-group makes the mean IQ of this group 100 and the standard deviation 15.

ENGLISH TESTS

Fig. 2 provides a comparison between the results of English and Inquiry pupils on one of the English tests. This is much more favourable to the Inquiry group. They are superior at the 95-percentile, about equal at the 50-percentile, and show a greater rate of increase of score at all levels.

ARITHMETIC TESTS

Fig. 3 gives the comparison for MHA 7. Perhaps the most striking feature is the small rate of increase in score for increasing

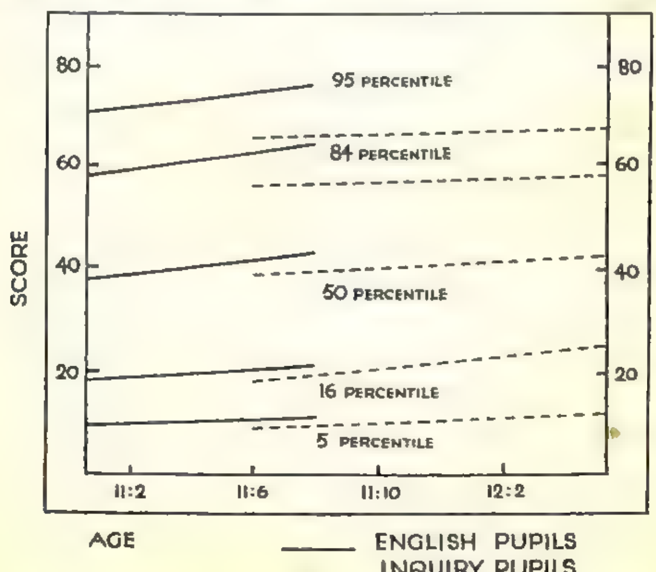

FIG. 3.—Arithmetic: Comparison between English and Inquiry pupils (MHA 7).

age at the high levels of ability in the Inquiry group. A possible explanation is that the clever children reach a uniform standard of achievement where variation in score is negligible. This may arise from the teacher's tendency to spend a great deal of time in bringing the duller children up to a certain level, thus giving little opportunity for the clever child to develop his ability to the full extent. With the English pupils the rate of increase at the high level of ability is greater than that at the lower levels, indicating that the clever children are making greater relative progress than the duller children. At all levels the English pupils show a performance superior to that of the Inquiry group.

The comparisons are all in respect of the December group, which is slightly poorer in quality than that presented in June. In general, however, the results for the June testing are in close agreement.

CHAPTER III

RELIABILITY OF THE TESTS AND EXAMINATIONS

Classification for Secondary Education on the Basis of Duplicate Qualifying Examinations

On the basis of the total marks the examiner was asked to divide the pupils into four categories:

A. Those fitted for a senior secondary course leading to the Senior Leaving Certificate.
B. Those fitted for a junior secondary course leading to the Junior Leaving Certificate.
C. Those fitted for a junior secondary course of a lower standard than B.
D. Those not fitted for promotion to any of the above types of course.

Category	Number of pupils in category in first examination	Distribution in second examination	
A	126	A	111
		B	14
		C	1
		D	0
B	82	A	25
		B	36
		C	21
		D	0
C	157	A	3
		B	22
		C	112
		D	20
D	135	A	0
		B	0
		C	28
		D	107
	500		500

To test the reliability of this classification a duplicate examination was set to the June group and the scripts of a representative

sample of 500 were corrected. The sample was selected by arranging the cards of the pupils who sat the first examination in order of merit and drawing every fourth. This gave nearly 500, the number being completed by drawing a few cards at random. The classification of pupils in the two examinations is shown on p. 13.

The agreement, particularly in the A and D categories, is much closer than one would expect from some of the data recently published on the unreliability of ordinary examinations, and it suggests that the correction and classification had been most carefully performed.

Reliability Coefficients of the Tests and Examinations

Another method of estimating reliability is to find the correlation between the two series of scores obtained when duplicate tests are applied to the same pupils. This is a *reliability coefficient*, and Dr Douglas M. McIntosh determined it for the Qualifying examination by using the results for the sample of 500. For the intelligence and scholastic tests he found the necessary data in the fact that nearly 1,400 pupils were common to the age-groups tested in December and June.

If we examine the raw coefficients in the first column of Table I perhaps the first point that strikes us is that the correlations for the ordinary examinations are unexpectedly high. It is also surprising to find that the reliability is higher for English than for Arithmetic, particularly in the examinations, where the English papers included an essay.

Change in order of marks from one examination or test to another may be the result of two causes: variation in the standard of performance of the individual pupils due to such factors as health, fatigue, suitability of the paper; and variation in the standard of marking of individual papers by the examiner. In the examinations both causes are operative; in the tests only the first, for the marking is done by a standardised system. For this reason we should expect the reliability of the tests to be higher than that of the examinations.

Reliability coefficients have to be interpreted with some statistical caution, since their size depends upon the length of the tests. In a sense it is unfair to compare the reliability coefficient of an intelligence test of 55 minutes with that of an examination lasting 3¼ hours. For a fair comparison we should reduce them

all to the same length, say an hour; and when this is done[1] we get the coefficients in the right-hand column of Table I. The tests now rise to the place which one would expect, and Q falls from the top of the list to the bottom. In this column Q is not Qe + Qa as in the first column, but an examination with half-hour papers in English and Arithmetic.

TABLE I

Reliability coefficients (raw and adjusted for length):
Tests and examinations

Test or examination	Reliability coefficient	Test or examination	Reliability coefficient if all had been of same length
Q *	+·933	E	+·953
E	+·931	A	+·934
Qe	+·920	I	+·921
I	+·898	Qe	+·847
A	+·877	Qa	+·845
Qa	+·864	(Q	+·811)

Qe, Qa, Q: N=500 I, E, A: N=1,350

* An explanation of the symbols used in the report will be found in the Glossary of Symbols in Appendix D.

Standard Errors of Measurement

A reliability coefficient depends also on the range of ability in the group for which it is calculated, and this factor must now be considered. We could correct the above coefficients for range, but for various reasons we use a different parameter, *the standard error of measurement*, which is small with a reliable test and large with an unreliable one. It has the advantage of being independent of the range. We can also calculate what it would be if the tests were all of the same length.

[1] The most convincing method would be to use tests and examinations which actually are of the same length; but, since this was not practicable, we made the reduction by a statistical procedure of which particulars will be found in J. P. Guilford, *Psychometric Methods*. New York: McGraw-Hill Book Company, Inc., 1936. P. 419.

TABLE II

Standard errors of measurement: Tests and examinations

Test or examination	Standard error of measurement * (percentage)	Test or examination	Standard error of measurement † (percentage)
E	4·41	E	3·61
A	5·15	A	3·77
Q	5·30	I	5·19
I	5·90	Qe	8·18
Qe	5·91	Q	8·91
Qa	8·65	Qa	9·23

* Comparable for range.
† Comparable both for range and for length of test.

The second set of percentages may be taken as giving the fairest comparison of the reliability of the various measuring instruments. They show that if the standardised tests had an equal chance with the examinations, through equalised range and length, they would be a much more reliable form of measurement. Their relative *validity* is, of course, another matter.

In both the tests and the examinations English is more reliable than Arithmetic; and the intelligence test is somewhat less reliable than the scholastic tests.

Since there is no variability in the marking of the tests the standard error of measurement of 3·61 per cent. in the English test appears to be due to other causes, particularly variation in the pupil's standard of performance. It is tempting, therefore, ignoring statistical rigour, to deduct this from the 8·18 per cent., which is the standard error of measurement of Qe; a very rough idea of the amount of the unreliability of the English examination which is due to variation in the marking is thus secured. This would indicate that nearly one-half of the unreliability of an ordinary examination is due to variation in the pupil's performance. If this conclusion were even approximately correct it would have alarming implications. In the report of the English International Institute Examinations Enquiry Committee[1] evidence is afforded of

[1] P. J. Hartog and E. C. Rhodes, *The Marks of Examiners.* London: Macmillan & Co., Ltd., 1936.

the remarkable variability in the marks given by different examiners to the same script. To this, it would appear, there has to be added an almost equally great variation arising from the fact that the single script is only a chance sample of one out of the pupil's possible performances in similar examinations.

The practical significance of a reliability coefficient of +·933 (the raw coefficient for Q) may be expressed in a less technical way as follows:

If a pupil's true mark in the Qualifying examination were 60 per cent., the chance that his actual mark would fall outside the range

55–65 per cent. is about 1 in 3,
50–70 per cent. is about 1 in 22,
45–75 per cent. is about 1 in 370.

While we have been at some pains to do justice to the newer forms of measuring device, the fact remains that in practice the Qualifying examination is much longer than the tests. The first set of percentages of Table II is therefore the one of most practical importance; and it shows that, as things are, a carefully set and administered examination can bear comparison, in regard to reliability, with standardised tests of intelligence and scholastic attainment. It should be mentioned, however, that the tests which we selected were shorter than some of those now used in Scotland.

CHAPTER IV

RELATIONS BETWEEN THE MEASURES USED AT THE QUALIFYING STAGE

The Tests and Examinations

In approaching the problem of selection for secondary education it is of interest to know the extent to which the various tests and examinations measure the same factors of ability. Clearly, if the correlation between two tests is $+1$, it does not matter which one we use, and there would be no point in using both. The problem is affected by the correlations between the measures in another way which the following example illustrates.

Suppose we had four Qualifying tests, A, B, C and D, of which the correlations with success in the secondary school (F) were $r_{AF} = +\cdot 7$, $r_{BF} = +\cdot 6$, $r_{CF} = +\cdot 5$, $r_{DF} = +\cdot 4$. If an Education Committee desired to use two of these tests for selection purposes, which pair should it select? Taken singly, A and B are the best tests, and there would be a temptation to choose them. It is, however, quite possible that the best pair to use would be the two worst, namely, C and D. This apparent anomaly is due to the fact that A and B may be measuring the same aspects of ability, and therefore duplicating each other's work, which would be indicated by a high correlation r_{AB}. On the other hand C and D may be measuring different aspects of the abilities on which success depends, and this would be shown by a low correlation r_{CD}. The choice, therefore, depends partly on the correlations of the tests with success and partly on their inter-correlations.

CORRELATIONS BETWEEN INTELLIGENCE AND THE OTHER TESTS AND EXAMINATIONS

We had two groups, the pupils presented in December 1935 and those presented in June 1936. We have therefore two

estimates in each case, but these were combined to give single correlations.[1]

TABLE III *

Correlations with intelligence: Tests and examinations (Qualifying and age-groups)

Correlation between I and	Qualifying group (N=2,900)	Age-group (N=5,600)
E	+·867	+·897
Qe	+·846	
A	+·759	+·801
Qa	+·726	
S	+·880	
Q	+·855	
Ts	+·791	
T	+·715	

* All the differences in correlation within the columns of this table are statistically significant. The method used in testing the significance of differences was to transform the correlation coefficients into z and calculate the standard error of the difference of z values.

In the correlation between intelligence and single tests for the Qualifying group the smallest z difference is ·074: SE=·026.

In the correlation between intelligence and combinations of English and Arithmetic for the Qualifying group the smallest z difference is ·101: SE=·026.

The z difference for the age-group correlations is ·356: SE=·019.

---

[1] In combining correlations the procedure used was as follows:

It was first ascertained whether the samples on which the separate correlations were determined could be regarded as coming from equally correlated populations.

The correlations were then transformed into z, where $z = \frac{1}{2} \log_e \frac{1+r}{1-r}$

The weighted average of the z values was found and re-transformed into r.

Certain special points in regard to combination of correlations are discussed in Chapter VII.

References: R. A. Fisher, *Statistical Methods for Research Workers.* Edinburgh: Oliver & Boyd, Ltd., 1936. P. 207.

P. R. Rider, *An Introduction to Modern Statistical Methods.* New York: John Wiley & Sons, Inc., 1939. P. 105.

The correlations are relatively high, a fact which is due partly to the wide range of ability in the groups, and partly, perhaps, to the great care with which the tests and examinations were administered.

The results show:

That intelligence (by which we mean simply whatever is measured by the intelligence tests which we used) is more highly correlated with English than with Arithmetic.

That intelligence is more highly correlated with the tests than with the examinations.

That intelligence is more highly correlated with the tests and examinations than with the teachers' estimates.

That the scaling of teachers' estimates [1] in such a way that they are comparable from school to school, by a method described in Chapter V, increases their correlation with intelligence.

The size of the correlations shows that the intelligence test is, to a considerable extent, covering the same mental ground as the other tests and examinations; and consequently we cannot expect a very large improvement in prediction when it is combined with, say, an English test. Also, in view of the fact that the correlation is higher with the tests than with the examinations, we must not be unprepared to find that a combination of an intelligence test with an examination is better for selection purposes than its combination with scholastic tests.

INTELLIGENCE SCORE AND INTELLIGENCE QUOTIENT

An important practical question is whether we should use the intelligence score or the intelligence quotient; and while we can settle this most conclusively from the follow-up correlations, some preliminary information may be obtained by a study of the correlations between the Qualifying data.

The correlation $r_{I\,IQ}$ is statistically spurious, but it is of some interest in this connection, in that it gives us an idea of the extent to which the ranking of pupils on intelligence score would be the

[1] The correlation with Ts was not, as in the case of the other coefficients in Table III, obtained by direct calculation. It was derived from the inter-correlations for the senior and junior secondary schools which were calculated for determining weightings, a correction for range being applied to make it comparable with the other coefficients in the Table. It is based on 2,160 subjects; and where the accuracy of the method could be tested, the agreement with the results of direct calculation was very close.

same as the ranking on intelligence quotient. The result is as high as $+\cdot 992$.[1]

The fact that this correlation is so high suggests that it would matter very little whether we used intelligence score or intelligence quotient in selection for secondary education. This does not follow: we must have a very high correlation indeed before we may draw such a conclusion.

This important point may be illustrated by actual figures. If we have two tests A and B such that $r_{AB} = +\cdot 992$, and if $r_{AF} = +\cdot 700$, what can we deduce as to r_{BF}? Without making assumptions which are normally unsound, it is not possible in such a case to determine r_{BF} exactly; all we can do is to fix limits between which its value must lie. When we make the calculation[2] we find that r_{BF} must lie between $+\cdot 604$ and $+\cdot 784$. Even if r_{AB} were as high as $+\cdot 999$, all we could say is that r_{BF} must lie between $+\cdot 668$ and $+\cdot 731$.

It does not follow, of course, that the correlations will differ as much as the above limits permit, and some results for the Qualifying group giving actual variations are shown on p. 22[3] ($N = 2{,}900$).

[1] This correlation was not calculated directly, but derived by means of the following formula from the correlation of intelligence score with age:

$$r_{\mathrm{IIQ}} = \frac{\dfrac{\sigma_{\mathrm{I}}}{M_{\mathrm{I}}} - r_{\mathrm{I}a}\dfrac{\sigma_a}{M_a}}{\sqrt{\dfrac{\sigma_{\mathrm{I}}^2}{M_{\mathrm{I}}^2} - 2r_{\mathrm{I}a}\dfrac{\sigma_{\mathrm{I}}\sigma_a}{M_{\mathrm{I}}M_a} + \dfrac{\sigma_a^2}{M_a^2}}}$$

M_{I} = mean of intelligence scores; σ_{I} = standard deviation of intelligence scores; a = age.

This formula assumes that, in expansions, powers of $\dfrac{d_{\mathrm{I}}}{M_{\mathrm{I}}}$ higher than the second may be neglected, d_{I} representing the deviation of an intelligence score from the mean.

Reference: G. Udny Yule and M. G. Kendall, *An Introduction to the Theory of Statistics.* London: Charles Griffin & Co., Ltd., 1937 Chapter xvi.

[2] The limits for r_{BF} are $r_{AF}r_{AB} \pm \sqrt{1 - r_{AB}^2 - r_{AF}^2 + r_{AB}^2 r_{AF}^2}$

Reference: G. Udny Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, p. 280.

[3] The correlations with IQ were obtained by the method described in the footnote on p. 20.

The differences between the correlations for IQ and I are barely significant statistically. The z differences are $\cdot 056$, $\cdot 050$ and $\cdot 061$: SE $= \cdot 028$.

Correlation with	IQ	I
S	+·892	+·880
Q	+·841	+·855
Ts	+·767	+·791
Average	+·833	+·842

The conclusion that it makes little difference whether we use intelligence score or intelligence quotient will be confirmed later by the follow-up correlations.

CORRELATIONS BETWEEN TESTS AND EXAMINATIONS

TABLE IV

Correlations between tests and examinations
(Qualifying group: N = 2,900)

Correlation between	r
E and Qe	+·866
A and Qa	+·810
S and Q	+·898

The correlations are again high, but even that of nearly ·9 between S and Q leaves ample room for differences in the selective value of the two types of measure. One interesting point that emerges is that the correlation between E and I is almost exactly the same as that between E and Qe.

CORRELATIONS BETWEEN ENGLISH AND ARITHMETIC

TABLE V

Correlations between English and Arithmetic
(Qualifying and age-groups)

Correlation between	Qualifying group (N = 2,900)	Age-group (N = 5,600)
Qe and A	+·737	
Qe and Qa	+·721	
E and A	+·708	+·766
E and Qa	+·664	

The higher correlation for the age-group is due to the wider range of ability.

The correlation between ability in English and ability in Arithmetic is a little above ·7; but it is interesting to notice how much its size depends on the nature of the measures used. Even if we take the highest correlation of ·766, it is clear that a knowledge of a pupil's standing in English gives very little guidance as to his standing in Arithmetic. From the graph given in fig. 27 [1] it can be seen that the error of prediction would be about 64 per cent. of what it would be by pure guess-work.

CORRELATIONS WITH TEACHERS' ESTIMATES

TABLE VI *

Correlations with teachers' estimates
(Qualifying group: N = 2,900)

Correlation between	r
Ts and Q	+·906
Ts and S	+·866
T and Q	+·774

* The first two correlations were obtained by the method described in the footnote on p. 20.

The differences are all significant statistically. The z difference between the first two is ·188: SE = ·030.

The correlation of nearly ·91 between Ts and Q is gratifying, but it gives no ground for the belief that the uniform examination could be replaced by the teachers' scaled estimates. Apart from the difficulty that the scaling is done on the basis of the examination marks, the correlation is still far too low.

Again we see the very significant rise in correlation that results from the proper scaling of the teachers' marks; and, as is to be expected, the scaled estimates correlate more highly with the examination than with the scholastic tests.

[1] P. 124.

EFFECT OF AGE ON THE CORRELATIONS

Correlations of the kind with which we are dealing may be considerably affected by lack of homogeneity of the groups. The Qualifying group is homogeneous in regard to the school class of the pupils, but very heterogeneous in regard to age. It is therefore necessary to ascertain whether the correlations are inflated by the fact that the data correlated have a common correlation with age.

One way in which we can do this is by comparing the correlations for the complete Qualifying group with those for a group which is reasonably homogeneous in regard to age; and for this purpose we chose the Qualifying pupils who fall within the age-group 11 years 6 months to 12 years 6 months. The correlations for the Qualifying group thus analysed into its component parts are given in Table VII.

TABLE VII *

Analysis of correlations for Qualifying group

Correlation between	Complete group 10.11–14.9 (N=2,900)	Middle 11.6–12.6 (N=2,170)	Extremes under 11.6 and over 12.6 (N=730)
I and E	+·867	+·928	+·878
I and A	+·759	+·759	+·773
I and Qe	+·846	+·842	+·843
I and Qa	+·726	+·727	+·719
I and Q	+·855	+·853	+·856
E and A	+·708	+·718	+·678
E and Qe	+·866	+·905	+·876
E and Qa	+·664	+·671	+·640
A and Qa	+·810	+·840	+·835
A and Qe	+·737	+·743	+·714
Qe and Qa	+·721	+·726	+·701
T and I	+·715	+·717	+·703
T and Q	+·774	+·780	+·756
Average	+·773	+·785	+·767

* The difference in average correlation between the complete group and the extremes is not significant statistically. The other average differences are significant. The z difference between the complete group and the middle is ·031: SE=·008. Between the middle and the extremes it is ·045: SE=·012.

The Table brings out the interesting result that in a Qualifying group the correlation is highest in the middle, and is considerably lower in the extremes; that is to say, it is highest where the age-scatter is least. The difference is no doubt due to the fact that in Qualifying classes the old children are usually dull and the young children clever. It is clear, too, that the correlations for a Qualifying group are much the same as those for a group which is homogeneous in regard both to age and to school class.

The above results can be shown from a different angle by using the correlations with age. We should expect a negative correlation in the extremes, but, in fact, *all* the correlations with age for the Qualifying group are negative. A small selection, calculated by Dr McIntosh, is given below:

Correlation[1] between age and	Middle (N=2,217)	Extremes (N=737)
Qe	−·086	−·129
Qa	−·112	−·129
Q	−·102	−·156
Average	−·100	−·138

We may conclude from this evidence that even in that part of a Qualifying group which is fairly homogeneous in regard to age the correlation with age is negative. This means that the older pupils are, on the whole, slightly poorer in ability and attainment than the younger.

USE OF PARTIAL CORRELATION METHODS

By the method of partial correlation we can find what the correlations would be if all the pupils were of the same age; and,

[1] All the correlations here are significantly different from zero. The probability that they could be obtained by random sampling from a population of zero correlation is of the order P=·0003.

In the few cases where it was necessary to test the significance of our correlations the formula used was

$$t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

The differences between the columns are not significant statistically. The z difference for the average is ·039: SE=·024.

References: R. A. Fisher, *Statistical Methods for Research Workers*, pp. 195 and 205.

P. R. Rider, *An Introduction to Modern Statistical Methods*, pp. 83 and 93.

as would be anticipated from the above results, the differences from the raw coefficients are extremely small. Out of five cases calculated by Dr McIntosh the largest difference was ·002.

EFFECT OF DIFFERENCES IN SCHOOL CLASS ON THE CORRELATIONS FOR THE AGE-GROUP

The Qualifying group is homogeneous in regard to school class but heterogeneous in age; the age-group, on the other hand, is relatively homogeneous in regard to age, but heterogeneous in regard to school class.[1] Our data provide us with an opportunity for making an analysis of the correlations for such a group; and, while this is not of much importance for our selection problem, it is of considerable interest in connection with the interpretation of correlation coefficients. In this case we shall count the pupils in the Qualifying class as the middle, and those in the classes above and below as the extremes.

TABLE VIII

Analysis of correlations for Qualifying and age-groups combined

Group	N	Average of three correlations *
Age-group outside Qualifying classes	3,414	+·844
Complete age-group	5,608	+·821
Combined age-group and Qualifying group	6,335	+·817
Qualifying group in age-group	2,194	+·802
Complete Qualifying group	2,921	+·778
Qualifying group outside age-group	727	+·776

* r_{IE}, r_{IA} and r_{EA}

While in the Qualifying group the correlation in the extremes is lower than that in the middle, or relatively homogeneous part, the opposite holds for the age-group.[2] The position is, therefore,

[1] *Cf.* pp. 2 and 3.

[2] The difference is statistically significant. The z difference is ·131: SE=·016.

that at the Qualifying stage in Scotland the inclusion of the old and young pupils in a group which is homogeneous in regard to school class lowers the correlations; but that the inclusion of pupils from higher and lower classes in a group which is homogeneous in regard to age raises the correlations. In this connection it should be mentioned that in the case of the age-group the standard deviation in the extremes is higher than the standard deviation in the middle, the opposite being true of the Qualifying group.

From the Table we see that the correlation for an age-grouping is higher than that for a class-grouping, the difference in correlation being about ·04.[1]

The Data from the Rating Cards

INDUSTRY

(In collaboration with Miss J. I. W. Robertson, M.A.)

The Qualifying teachers were asked to grade all pupils in regard to industry, using a five-point scale in which the significance of the letters was as follows:

- A Exceptionally high.
- B High—decidedly above the average.
- C Average.
- D Low—decidedly below the average.
- E Exceptionally low.

The leaflet of instructions gave the percentages of pupils who would fall into the various categories in a *normal* class.

Using the June group, Miss Robertson calculated correlations between these industry gradings and intelligence quotient, Qualifying examination marks, and health. She dealt with boys and girls separately.

All correlations which we have calculated for data on five-point scales have been corrected for broad grouping.[2]

[1] The difference is statistically significant. The z difference is ·119: SE=·013.

[2] References: K. Pearson, "On the Measurement of the Influence of Broad Categories upon Correlation," *Biometrika*, vol. ix, 1913, p. 116.
K. J. Holzinger, *Statistical Methods for Students in Education.* New York: Ginn & Co., 1928. Chapter xiv.

TABLE IX *

Industry: Correlations with IQ, Q, H
(June Qualifying group)

Correlation between industry and	Boys		Girls		All	
	N	r	N	r	N	r
IQ	953	+·598	987	+·647	1,940	+·622
Q	931	+·568	956	+·664	1,887	+·616
H	736	+·112	763	+·140	1,499	+·126

* The correlation with health is significant statistically, P being far below ·01. As far as the differences between boys and girls are concerned, the difference between the correlations with Q is significant; z difference is ·155: SE = ·045. In the other two cases the differences are not significant, though that for IQ is nearly so; z difference is ·080: SE = ·045.

It may be that there is a tendency for teachers to ascribe a high degree of industry to a child who is doing well in class, while his success may be due rather to superior ability or home encouragement. On the whole, however, we felt that the industry gradings were much more reliable than those for interest, which we shall discuss presently.

Taking them at their face value, they show a correlation of about ·6 both with intelligence and scholastic attainment and only a small positive correlation with health. From our point of view the results offer hope that industry may be of some value in forecasting success, since its correlation with the other measures is such as to suggest that it may cover a part of the field which they leave untouched. Perhaps the result that gives most food for thought, however, is the size of the correlation with IQ.

It is interesting, too, to notice that the correlations for the girls are all higher than those for the boys, though the difference is significant statistically only with Q. Some further data on the question of the difference in industry between boys and girls are given in Table X.

The Table gives the percentages falling into the various categories. They are comparable as between industry and Qualifying examina-

tion marks, for the latter distribution was cut in such a way that the dividing lines between the categories are at the same σ distance from the mean in both cases. The means at the foot of the columns are found by giving five points for an A grading, four points for a B and so on, and may be taken as giving a rough comparative measure of the quality of the groups.

TABLE X

Industry: Comparison of boys and girls
(June Qualifying group)

Category	Boys (N=931)		Girls (N=956)	
	Ind.	Q	Ind.	Q
A	3·0	2·0	3·0	2·9
B	26·5	30·0	26·8	30·5
C	53·6	48·6	56·8	50·4
D	15·0	16·9	11·5	14·8
E	1·8	2·6	1·9	1·2
Mean	3·14	3·12	3·18	3·19

The differences between boys and girls are statistically significant both in industry and in Qualifying examination marks, the former difference being nearly four times its standard error. On the whole, therefore, the results support the view that girls are more industrious than boys, though the difference is not very great. A closer study of the table shows where the difference mainly lies. The good boys are nearly as industrious as the good girls, but the boys of poorer ability are less industrious than girls of the same type.

PUPILS' INTERESTS

(In collaboration with Mr G. A. Crabb, B.Sc.)

Predictive Value of Teachers' Interest Gradings

With a view to exploring the value, in guidance and selection, of information as to the direction of the pupils' interests, we asked

the primary teachers to grade each Qualifying pupil on a five-point scale in regard to interest and attainment in the three main groups of school subjects—humanistic, realistic and practical. The meanings of the letters were the same as in industry.

Mr Crabb made a study of these interest gradings; and, while his results did not encourage us to proceed with the statistical working-up of this part of our data, he continued his work on the subject of children's interests until his entry on war service. The following is an account of the part of his investigations which bears upon the present Inquiry: his further results he will, no doubt, publish independently.

The group chosen was that presented in June, which showed a good distribution of pupils among the five categories. When, however, the gradings were compared with those for attainment there was ground for suspicion that the agreement was greater than was probable on the basis of a true assessment of interest. Mr Crabb made the comparison in two ways: by finding the numbers of pupils who came into the same category both for interest and for attainment, and by finding the correlation, corrected for broad grouping, between the interest and attainment gradings. By the first method he found that no less than 84 per cent. had the same grading in both. The correlation between the interest and attainment gradings in humanistic subjects was $+\cdot 94$.

Mr Crabb planned an interesting subsidiary experiment to find how far the teachers' estimates of interest corresponded with the children's own ideas on the subject. A new group was formed, consisting of the Qualifying pupils in five representative schools. These pupils were graded by the teachers in exactly the same way as the Inquiry group had been. They were then given a simple Interests questionnaire.

Sixteen school subjects or parts of subjects (for example, Poetry, Composition, Spelling) were listed, and the pupil was asked to indicate by underlining whether he liked the subject, disliked it, or was uncertain whether he liked or disliked it. For purposes of scoring, a *like* was credited with three marks, a *not sure* by two marks, and a *dislike* by one mark. The subject selected as *favourite* was given an extra mark.

The school subjects were then divided into humanistic, realistic and practical, so that the pupil's score in each group could be obtained. By cutting the distribution of scores in such a way that the border-lines were at the same σ distance from the mean both

in the teachers' gradings and in the pupils' own scores, a comparable five-point grading, made by the pupils themselves, was secured.

TABLE XI

Interest: Comparison of gradings made by teachers with those made by pupils (special experimental group: N=227)

Group of subjects	Correlation between teachers' and pupils' gradings	Percentages of pupils in same category in both gradings
Humanistic . . .	+·26	34·8
Realistic . . .	+·46	44·0
Practical . . .	+·13	39·6

The average correlation between the teachers' gradings in interest and attainment was +·95, which is very close to that obtained for the Inquiry group. The average correlation between the pupils' gradings of their own interests and the teachers' gradings for attainment was +·40.

These results show that there is great disparity between the teachers' estimates of children's interests and the children's own ideas; and they tend to confirm the suspicion that the teachers' gradings were largely influenced by the pupils' attainments. That such may be the case is no reflection upon the care with which the teachers made their estimates: the fault lay in our technique, which set them an impossible task. We decided, however, that instead of calculating follow-up correlations for the interest data on the rating cards, Mr Crabb should follow up the pupils in his experimental group with a view to comparing the prognostic value of the teachers' and pupils' gradings. This project had to be abandoned on account of his entry upon war service.

Pupils' Order of Preference of School Subjects

The method of scoring used by Mr Crabb enabled him to determine comparative interest scores for the various school subjects, and, though the number of pupils was rather small, fig. 4 is presented as a contribution to the mounting evidence on this topic. It gives the order of preference for a group of pupils at the Qualifying stage in a Scottish industrial City, and shows also the range of variation in the placing of the subjects by individual

pupils. The centre points give the mean placing of the subjects, and the horizontal lines cover a range of $\pm 3\sigma$.

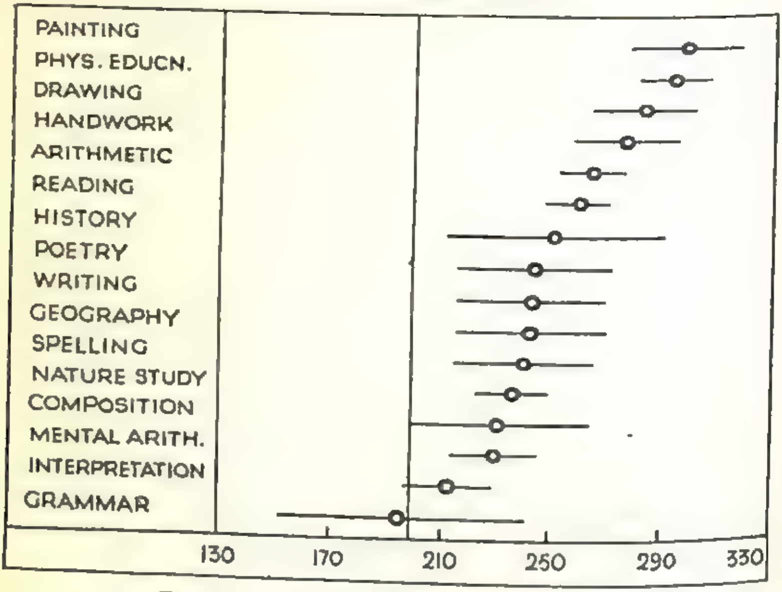

FIG. 4.—Pupils' preferences in school subjects.

The vertical line between 170 and 210 divides subjects where the number of *likes* is greater than the number of *dislikes* from those where the reverse is the case. It could hardly be assumed that this gives a border-line between absolute liking and dislike; one suspects that the child gives his judgment with the reservation 'as far as school subjects go.' Still, the results confirm the view that Grammar is a distasteful subject to the majority of pupils at this age. Yet even here, as a cautious boy put it, 'everything depends on the teacher'; in one class the gradings for Grammar were: *like*, 60 per cent., *not sure*, 40 per cent., *dislike*, 0 per cent., *favourite subject*, 30 per cent.

Comparison of Interests of Boys and Girls

The number of pupils in this experiment was too small to yield reliable results, but, generally speaking, they give no ground for the belief that there is any significant difference between the educational interests of boys and girls at this stage. There were small differences in the order of preference of school subjects but only one or two stood the test of statistical significance. Nor, as is shown in Table XII, is there any statistically significant difference between boys and girls in the three main groups of subjects.

TABLE XII

Interest: Comparison of mean scores of boys and girls in humanistic, realistic and practical subjects
(special experimental group: N = 227)

Group of subjects	Boys (N = 116)	Girls (N = 111)
Humanistic . . .	2·31	2·39
Realistic . . .	2·46	2·42
Practical . . .	2·75	2·83

As might be expected, both groups show the greatest preference for the practical subjects, and a slightly greater preference for the realistic than for the humanistic subjects.

CHAPTER V

TEACHERS' ESTIMATES

(In collaboration with Mr John T. Bain, M.A., B.Sc.)

Variation in Standard from School to School

On a preliminary survey the teachers' estimates appeared to have the following characteristics:

They were generally too high.

They tended to over-estimate the attainments of the weak pupils and under-estimate those of the good pupils.

The standard varied considerably from school to school.

The first of these features is not very serious from the point of view of follow-up correlations. The second decreases the weight given to teachers' estimates when they are combined with other measures having a larger standard deviation. The third is the most important, for it would lower the correlations considerably.[1]

To ascertain the extent of the variation in standard, Mr Bain analysed the data for a representative group of twelve schools comprising about 850 pupils in all.

At the moment we are concerned only with the columns in Tables XIII and XIV which are headed Te, Qe, Ta and Qa. They show that in English the teachers' estimates are, on the average, 15 per cent. higher than the scores made by the same pupils in the uniform examination; in Arithmetic they are 25 per cent. higher.

In English the teachers' estimates are much less widely scattered than the examination marks, the average standard deviations being 12·3 and 18·4 respectively; in Arithmetic the difference in scatter is very much less.

The variations in standard from school to school are very striking. In English the teachers' estimates for School 4 are on the average 1·9 per cent. above the mean mark scored in the examination; in School 9 they are 30·2 per cent. higher. The difference in standard of marking between these two schools is

[1] *Cf.* Chapter VII.

therefore more than 28 per cent. In Arithmetic the maximum difference in standard is 22 per cent.

TABLE XIII *

Teachers' estimates in English: Comparison with marks in English scored by the same pupils in the Qualifying examination

School	Te		Qe		Tes		r_{QeTe}
	M	σ	M	σ	M	σ	
1	64·7	14·7	56·6	19·9	56·2	18·9	+·94
2	60·4	12·3	42·1	18·2	37·8	15·5	+·72
3	64·9	10·7	62·0	17·2	60·2	15·8	+·79
4	63·9	10·0	62·0	16·3	60·7	16·1	+·89
5	67·5	11·1	60·7	16·7	57·8	16·9	+·87
6	65·3	11·0	52·2	18·4	52·3	17·1	+·82
7	74·7	10·8	58·1	18·9	54·2	18·4	+·85
8	65·2	15·5	45·8	22·1	47·2	21·1	+·91
9	76·1	13·2	45·9	16·5	45·8	18·1	+·82
10	61·1	18·3	38·3	19·5	37·1	17·6	+·83
11	60·5	7·9	39·5	15·4	36·4	12·4	+·60
12	67·2	9·6	44·3	18·1	44·7	15·1	+·76
Average . .	66·1	12·3	51·1	18·4	49·6	17·1	+·82
Slumped schools	66·1	13·6	51·1	20·3	49·6	19·1	+·77

* The numbers of pupils have been omitted to avoid the possibility of the identification of particular schools. The averages at the foot of the columns are weighted averages. Marks are in percentages.

It is thus confirmed that teachers' estimates have the three features mentioned above. In regard to the first it has to be kept in mind that if the estimates are higher than the examination marks it does not follow that it is the teachers' standard that is wrong, while the difference in scatter may be partly due to the practice of basing estimates on the average of a large number of class tests. Such averages will normally have a smaller scatter than the marks in a single examination.[1]

[1] Since $\sigma^2_{(X+Y+Z+\ldots)} = \Sigma\sigma^2_X + 2\Sigma r_{XY}\sigma_X\sigma_Y$

TABLE XIV

Teachers' estimates in Arithmetic: Comparison with marks in Arithmetic scored by the same pupils in the Qualifying examination

School	Ta		Qa		Tas		r_{QaTa}
	M	σ	M	σ	M	σ	
1	69·0	24·3	52·4	20·9	52·8	22·0	+·88
2	69·5	16·5	35·9	16·0	37·0	17·2	+·75
3	71·3	13·4	51·6	17·3	51·2	17·6	+·81
4	72·8	16·5	58·5	16·3	57·6	15·8	+·78
5	72·0	19·4	50·4	18·1	47·6	16·6	+·80
6	68·5	11·1	55·5	15·7	53·7	14·0	+·70
7	79·1	13·8	51·7	18·7	49·6	15·0	+·75
8	74·5	20·3	40·3	17·2	42·6	20·8	+·82
9	77·4	22·9	42·5	14·7	42·5	16·4	+·76
10	62·6	22·8	37·0	22·0	38·8	24·8	+·82
11	58·6	10·7	26·6	13·3	23·0	12·3	+·74
12	65·8	18·1	38·1	19·9	40·1	21·6	+·61
Average . .	70·4	17·6	45·6	17·9	45·4	18·3	+·77
Slumped schools	70·4	18·8	45·6	19·9	45·4	20·1	+·73

Comparison of Variations in Standard for English and Arithmetic

The question is whether a school that marks leniently in English marks leniently to the same extent in Arithmetic, and whether a school whose marks are widely scattered in English shows a similar wide scatter in Arithmetic. These are, as we shall see, important points in connection with the scaling of teachers' estimates.

In making the comparison it is difficult to obtain figures which can be readily compared in view of the fact that the teachers' marks in Arithmetic are on the average 24·8 per cent. higher than those in the examination, while in English the difference is only 15 per cent. We have therefore added 9·8 per cent. to the Qa

average, and made a corresponding correction in the standard deviations.[1]

TABLE XV

Teachers' estimates in English and Arithmetic: Comparison of variations in standard of marking

School	M		σ	
	Te – Qe	Ta – Qa	Te – Qe	Ta – Qa
1	+ 8·1	+ 6·8	+ ·6	+3·4
2	+18·3	+23·8	− ·1	+ ·5
3	+ 2·9	+ 9·9	− ·7	−3·9
4	+ 1·9	+ 4·5	− ·5	+ ·2
5	+ 6·8	+11·8	+ ·2	+1·3
6	+13·1	+ 3·2	−1·6	−4·6
7	+16·6	+17·6	−2·3	−4·9
8	+19·4	+24·4	− ·8	+3·1
9	+30·2	+25·1	+2·5	+8·2
10	+22·8	+15·8	+4·6	+ ·8
11	+21·0	+22·2	−1·7	−2·6
12	+22·9	+17·9	−2·7	−1·8

The answer to our question is, therefore, that we should make serious errors if we assumed that a school which marks leniently in English marks equally leniently in Arithmetic, or that a school whose marks are widely scattered in English shows a correspondingly wide scatter in Arithmetic.

Correspondence between Orders of Merit in Teachers' Estimates and Uniform Examination

Mr Bain calculated correlations between teachers' estimates and Qualifying examination marks for each school, for English and Arithmetic separately. These are given in the last columns of Tables XIII and XIV, from which it will be seen that the average

[1] It should be noted that the standard deviations for the slumped schools in Tables XIII and XIV have a different meaning from the average standard deviations. For this correction we use the average standard deviations.

correlations for English and Arithmetic are +·82 and +·77 respectively. The difference is statistically significant, so we may conclude that there is a closer correspondence in English than in Arithmetic.[1] In neither subject, however, is the correspondence as close as we might expect in marks which purport to measure the same abilities.

Methods of Reducing Teachers' Estimates from different Schools to the same Standard

The above results show that teachers' estimates, as they stand, should not be used in selection and guidance. Before using them some means must be found for making them comparable in standard, and there are two ways in which this improvement might be effected.

guidance to teachers as to standard for estimates

The first line of attack is to devise ways in which teachers could be helped to use a uniform objective standard of marking, but all that we could attempt in this direction was a preliminary investigation of one of the possible methods.[2] We asked the teachers to give estimates on a five-point scale where each category had a definite significance, the percentages of a normal class that would fall into each category being indicated in the leaflet of instructions. The gradings were made for attainment only, taking no account of the pupils' ages.[3]

Mr Bain investigated the question whether these letter estimates were on a more uniform standard than the percentage marks. He took the five-point gradings in attainment in humanistic and realistic subjects, determined the points on a percentage scale which corresponded to the means of various categories, and thus found a rough measure of the means and standard deviations for the various schools. These are shown in Table XVI.

We cannot attach much importance to the fact that the averages of the means on the five-point scale are much closer to the corre-

[1] z difference is ·136: SE = ·049.

[2] A useful discussion on this subject will be found in the report of the Swedish Committee of the International Institute Examinations Inquiry—F. Wigforss, *The Entrance Examination in view of Later School Performances.* Stockholm: P. A. Norstedt & Sōners Förlag, 1937. P 110.

[3] For the significance of the letters see p. 27.

TABLE XVI

Teachers' estimates on five-point scale: Improvement in uniformity of standard

School	English			Arithmetic		
	Te (marks)	Te (5-point scale)	Qe	Ta (marks)	Ta (5-point scale)	Qa
1	64·7	55·2	56·6	69·0	58·6	52·4
2	60·4	45·8	42·1	69·5	49·8	35·9
3	64·9	44·2	62·0	71·3	51·4	51·6
4	63·9	54·8	62·0	72·8	60·0	58·5
5	67·5	53·1	60·7	72·0	52·6	50·4
6	65·3	51·7	52·2	68·5	52·7	55·5
7	74·7	51·4	58·1	79·1	51·4	51·7
8	65·2	50·9	45·8	74·5	53·0	40·3
9	76·1	50·4	45·9	77·4	50·9	42·5
10	61·1	46·6	38·3	62·6	44·2	37·0
11	60·5	46·5	39·5	58·6	45·0	26·6
12	67·2	48·2	44·3	65·8	50·9	38·1
Average M	66·1	49·6	51·1	70·4	51·7	45·6
Average σ	12·3	14·3	20·3	17·6	15·4	19·9

sponding figures for the uniform examination, for this depends on the statistical adjustment. To find the improvement in standard of marking, both forms of estimate were reduced to the same means as the corresponding examination marks and the average differences calculated. These are given in Table XVII,[1] from which it will be seen that the improvement by the use of the five-point scale is 1·1 per cent. That this is not sufficient is confirmed by a scrutiny of the results for the individual schools. In English, with the five-point scale, School 3 under-estimates by 17·8 per cent., while School 10 over-estimates by 8·3 per cent.

Mr James C. Kidd, M.A., B.A., investigated the correlations of the five-point estimates with later success in the senior secondary

[1] P. 43.

courses, and found that they were all lower than the corresponding correlations for the percentage estimates.

We conclude that a considerable improvement in technique would be necessary before the five-point method would yield results that could profitably be used in practice.

THE SCALING OF TEACHERS' MARKS

The second method is to take the teachers' estimates as they are and to reduce them to comparability by statistical procedure. For this purpose it is not sufficient to use standard scores, since this would assume that the pupils of all the schools had the same mean level of attainment. Clearly we must transform the teachers' marks in such a way that the mean and standard deviation of the transformed marks for any school will be the same as the mean and standard deviation of the marks made by the pupils of that school in a uniform test or examination in the same subject. This can be done by the transformation [1]

$$Ts = M_Q + \frac{\sigma_Q}{\sigma_T}(T - M_T)$$

and this, in our view, is the soundest method to adopt. It is, moreover, simple enough in practice. Since the equation is linear, the constants are easily calculated and a transformation Table can be quickly made.

It has, unfortunately, one disadvantage: in certain cases the transformed marks fall outside the limits of the percentage scale. The statistician would readily reconcile himself to this, but it is doubtful whether an Education Committee would view with favour any method which ascribes a negative mark to a pupil. Such negative marks are, of course, not so absurd as they seem at first sight. They do not mean that the pupil has less than a zero knowledge of the subject, merely that he is a certain distance below some arbitrarily chosen point of reference. Apart from the possible misunderstandings, however, there is a practical difficulty in the use of negative scores in that they are a fruitful source of error.

METHOD USED IN THE INQUIRY

We experimented with various methods of overcoming the above difficulty, keeping in mind the fact that the system used

[1] K. J. Holzinger, *Statistical Methods for Students in Education*, p. 121.

should be one that could be carried out quickly by the ordinary clerical staff of an Education Office; and in the end the following procedure was adopted. We took the teachers' estimates and the examination marks for a school and equated the quartiles and ends of the ranges, giving a graph of the type shown in fig. 5. A separate graph must of course be made for each school. If the raw mark falls at P we read the transformed mark at Q.

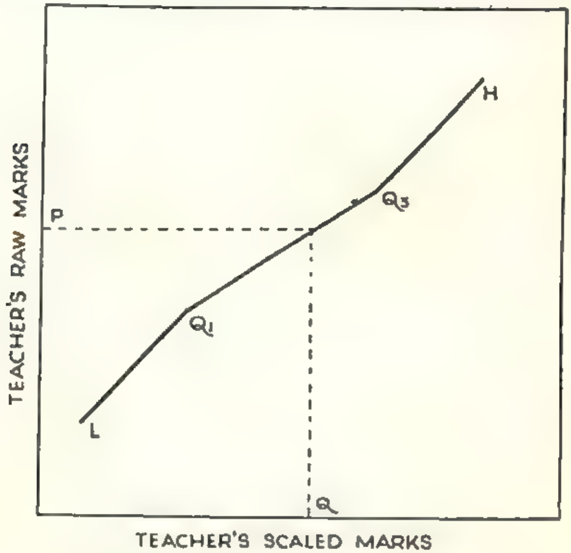

FIG. 5.—Graph for scaling teachers' marks.

While the equation of the quartiles is not open to serious statistical objection, it was with some hesitation that we decided to use the highest and lowest scores. We must have two fixed points outside the quartile range $Q_1 - Q_3$ if we wish to avoid negative scores, yet in using the ends of the ranges we felt that we were too much dependent upon accident: a pupil who had made an exceptionally low score in the examination would pull down the marks of all pupils below the lower quartile. The procedure was, however, given a very searching test through its application to the whole Qualifying group, and our justification for offering it as a practical solution of the difficulty is that with no pupil did it appear to operate unfairly.

Difficulties in the Use of Scaling Systems

The main practical difficulty is that *no* scaling system can work satisfactorily with small numbers, and these may occur in two cases:

In country districts the number of Qualifying pupils in any one school may be very small.

In urban systems where the three-track plan is used the numbers in the lowest section may be small, even in a school of considerable size.

Of these we had to contend only with the second, and the difficulty did not prove to be serious. We ascertained what procedure had been used in making the estimates; and fortunately most schools had given the marks for the different sections on the same standard. In a small number of schools the C section pupils were marked more leniently, and the estimates had to be scaled separately.

The difficulty arising from the three-track plan can be overcome by requiring the teachers to give their estimates on the same standard for all pupils. For the small rural school the system of scaling is quite impracticable; and, unless uniformity of standard can be secured in other ways, the use of teachers' estimates should be abandoned.

A third possibility of breakdown in the scaling system would not arise if our final recommendations are accepted. In certain Areas a special examination is held for admission to secondary schools or for the award of bursaries, and it may happen that the number of candidates from a particular school is very small. We believe, however, that any sound system of guidance and selection at the Qualifying stage demands the examination of the complete Qualifying group, by which we mean all the pupils in the Qualifying classes.

Improvement effected by Scaling

The improvement effected by scaling can be inferred from Tables XIII and XIV, where the school means of the teachers' transformed estimates are given. The standards for the schools are now closely in conformity with the marks scored by their pupils in the uniform examination, and the scatter is approximately that given by the examination. If we make all the average means the same, we obtain a fair comparison of the average

difference between the school means and the examination means for the three types of estimates with which we have experimented.

TABLE XVII

Teachers' estimates: Average difference between school means and means of Qualifying examination marks for various types of estimate

Type of estimate	Average difference (per cent.)	
	English	Arithmetic
Unscaled percentages . . .	7·4	6·7
Five-point scale	6·3	5·6
Scaled percentages . . .	1·5	1·6

Perhaps the most important result of the scaling is that the *large* differences in standard have entirely disappeared.

Separate Scaling for English and Arithmetic

So far we have examined the procedure where English was scaled on English and Arithmetic on Arithmetic. This involves a good deal of labour, which could be halved if we simply scaled the totals for the two subjects. To consider whether this would be legitimate or desirable we have to introduce a somewhat cumbrous notation:

Let Te = teachers' raw estimates in English,
Tes = teachers' estimates in English scaled on English,
and similarly for Ta and Tas,
T = teachers' raw estimates in English + Arithmetic.
T is therefore equal to Te + Ta.

We can arrive at the teachers' scaled estimates for English + Arithmetic in two ways. Te may be scaled on Qe, giving Tes, and Ta on Qa, giving Tas. This method gives Ts = Tes + Tas. We may, on the other hand, scale Te + Ta on Qe + Qa, obtaining a result which we shall denote by $T's$.

Our problem is to ascertain whether, for purposes of guidance and selection at the Qualifying stage, it would be legitimate or desirable to use $T's$ instead of Ts. This is a question to which

the most satisfactory answer would have been obtained by using both methods of scaling and by calculating follow-up correlations for the resulting scaled estimates. This was not done, partly because of the labour involved, and partly because we have other reasons, which we shall state presently, for preferring Ts to Ts′. Without actually scaling we can roughly deduce the predictive value of Ts′ as compared with Ts. The method used has a somewhat mathematical appearance; and as it is too long for a footnote it has been put into an indented section which the general reader is advised to omit.

Since the method of scaling adopted is an approximation to the exact way of making the mean and standard deviation of the teachers' estimates equal to the mean and standard deviation of the marks in the uniform examination, we may take it that

$$\text{Tes} = M_{\text{Qe}} + \frac{\sigma_{\text{Qe}}}{\sigma_{\text{Te}}}(\text{Te} - M_{\text{Te}})$$

It can then be shown that, for a single school,

$$r_{\text{TsTs}'} = \frac{\sigma_{\text{Qe}}\sigma_{\text{Te}} + \sigma_{\text{Qa}}\sigma_{\text{Ta}} + (\sigma_{\text{Qe}}\sigma_{\text{Ta}} + \sigma_{\text{Qa}}\sigma_{\text{Te}})r_{\text{TeTa}}}{\sqrt{\{\sigma^2_{\text{Te}} + \sigma^2_{\text{Ta}} + 2r_{\text{TeTa}}\sigma_{\text{Te}}\sigma_{\text{Ta}}\}}\sqrt{\{\sigma^2_{\text{Qe}} + \sigma^2_{\text{Qa}} + 2r_{\text{TeTa}}\sigma_{\text{Qe}}\sigma_{\text{Qa}}\}}}$$

which would not normally be equal to unity.

For one of the schools in Mr Bain's sample $r_{\text{TsTs}'} = +\cdot 96$. This is a fairly high correlation, but, as we have seen, we cannot without other evidence say that it would not matter which we used for prognostic purposes.

It can also be shown that

$$r_{\text{Ts}'\text{F}} = \frac{\sigma_{\text{Te}}r_{\text{FTe}} + \sigma_{\text{Ta}}r_{\text{FTa}}}{\sqrt{\{\sigma^2_{\text{Te}} + \sigma^2_{\text{Ta}} + 2r_{\text{TeTa}}\sigma_{\text{Te}}\sigma_{\text{Ta}}\}}}$$

and

$$r_{\text{TsF}} = \frac{\sigma_{\text{Qe}}r_{\text{FTe}} + \sigma_{\text{Qa}}r_{\text{FTa}}}{\sqrt{\{\sigma^2_{\text{Qe}} + \sigma^2_{\text{Qa}} + 2r_{\text{TeTa}}\sigma_{\text{Qe}}\sigma_{\text{Qa}}\}}}$$

From our other data we know the limits between which r_{FTe}, r_{FTa} and r_{TeTa} are likely to vary. We can, therefore, make rapid comparisons using the schools in Mr Bain's sample and experimenting with different values of the correlations. Such experiments are not a very rigorous way of settling the question, but they lead to the following conclusions:

In the great majority of cases r_{TsF} would be higher than r'_{TsF} but in exceptional circumstances r'_{TsF} could be the greater.

For a single school of reasonable size r_{TsF} would rarely exceed r'_{TsF} by more than ·03: the difference would normally be about ·02.

Our conclusion is that the correlation with later success of teachers' estimates scaled separately for English and Arithmetic would be higher, probably by about ·02, than that for estimates scaled for English and Arithmetic together. A stronger case for the first method can be made out on other grounds, of which the most compelling is the need for satisfactory separate estimates of the pupil's ability in the two subjects.

Scaling Teachers' Estimates on Intelligence Test Results

The educational importance of this problem is much greater than at first sight appears. It lies in the fact that, if we could legitimately scale teachers' estimates on the intelligence test, and if a combination of an intelligence test and teachers' scaled estimates gave a high correlation with later success, we could avoid all the educational evils of uniform external examinations at the Qualifying stage.

At this point we can deal only with the question whether the method would be legitimate, and we take first the simpler problem of the scaling of the total marks in English + Arithmetic. For this purpose we use as our standard of comparison the teachers' marks scaled by the soundest possible method, that is, on Qualifying examination scores. With them we must compare the teachers' estimates as scaled, say, on IQ, which we shall call Ts'I.

Mr Bain investigated this point with his sample of twelve schools.

Glancing along the rows of Table XVIII we can see that scaling on IQ makes a considerable improvement, but not so great as that made by scaling on Q. Using the method previously adopted we compare the improvements in Table XIX. This is a Table which we present with considerable regret. It shows that the method of scaling upon IQ cannot be regarded as sound. Even in the small sample of twelve schools injustice to the extent of over 12 per cent. could be done to considerable numbers of pupils, and it seems unfair to use a system which has a known error of this size.

TABLE XVIII

Teachers' estimates scaled on Q and on IQ

School	Q	Ts	T'sI	T
1	55·6	55·1	54·9	67·3
2	39·8	40·0	44·0	64·0
3	57·5	56·9	62·0	67·2
4	60·4	58·7	56·0	68·6
5	56·4	55·8	54·8	69·0
6	53·8	52·9	52·0	66·5
7	55·6	53·6	54·1	75·4
8	43·6	44·7	51·6	69·5
9	44·1	43·8	49·9	77·0
10	37·8	37·9	43·2	61·0
11	34·6	34·0	41·5	59·0
12	41·7	44·4	46·8	66·2
Average M . .	48·9	48·7	51·3	67·7
Average σ . .	16·8	16·6	15·1	14·0

TABLE XIX

Teachers' estimates scaled on Q and on IQ: Average difference between school means and means of Qualifying examination marks

Type of estimate	Average difference per cent.	Maximum error * per cent.
T	6·9	24·7
T'sI	3·8	12·4
Ts	·9	4·7

* Maximum error = sum of largest positive and negative differences.

Mr Bain calculated the following correlations with teachers' estimates scaled on the Qualifying examination and on the intelligence test for his group of 842 pupils:

Correlation of Q and T (average) +·836
Q and T (slumped) +·787

Q and T'_s (slumped) +·883
Q and T'_{s1} (slumped) +·857

T'_s and T'_{s1} (slumped) +·971

The most interesting of these correlations is the last. Despite the fact that the teachers' estimates scaled in these two ways correlate to the extent of ·971, we should, as we have seen, do an injustice amounting to as much as 12 per cent. to certain pupils if we used the method of scaling on IQ. The middle pair show that the marks scaled on the examination correlate more highly with the examination than the same marks scaled on IQ. The first pair give another example of the way in which correlations are lowered when we slump together schools where there is a variation in standard of marking.

Another disadvantage of the intelligence test as a scaling basis is that we cannot obtain reliable estimates in English and Arithmetic separately without assuming that the relative positions of the school means are the same in both subjects. The unsoundness of this assumption is clearly brought out by Table XV and by the following correlations between school means for the June Qualifying group, comprising 36 schools:

	Correlation between school means
IQ and Qe	+·88
IQ and Qa	+·86
Qe and Qa	+·89

These are far from high enough to justify a scaling of English or Arithmetic on IQ.

Conclusions

We conclude, therefore, that the only satisfactory procedure is to scale the estimates separately for English and Arithmetic, and that the estimates should be scaled upon the results of a uniform examination.

Steps should at the same time be taken to ensure that all the teachers make their estimates on the same basis. The instructions should make it clear whether the estimates are to be measures of the pupils' attainments at the time of examination, or whether they are to take account of evidence of progressive improvement or deterioration; also whether allowance is to be made for ability, age and personal qualities.

CHAPTER VI

COMPARABILITY OF SUCCESS MARKS IN THE SECONDARY COURSES

The Measure of Success

To compare the prognostic value of the various measures used at the Qualifying stage we require a criterion by which they can be judged, and this should, strictly speaking, be the child's *ability to profit* by the secondary course. For scientific investigation, however, we required one that could be expressed in exact quantitative terms; and we used the child's success. The two, we grant, are not quite the same: a child may profit much more from a secondary course than his success mark would indicate. Yet, from the point of view of the child, the parent and the Education Committee, it is success that is the more important. Moreover, the pupil who consistently scores low marks in his secondary school examinations has much on the debit side of his 'ability to profit' account, no matter how much he has on the credit side in the way of response to cultural stimulus and social contacts.

Accepting success as our criterion, we have still to determine what we are to take as its measure; and after careful consideration we decided to use the weighted average of the marks in all the subjects of the secondary course.

Doubt as to the soundness of this procedure arose only once, when, in a discussion with the head masters of the senior secondary schools, the suggestion was made that it would be better to discard such subjects as Art, Handwork, Music and Physical Education, and take an average merely of the important and academically reputable subjects like English, Latin, French and Mathematics. It is, of course, quite true that from many points of view the subjects which matter most are those in which the pupil will aim at passes on the higher grade of the Leaving Certificate examination, and which he will require for such purposes as university entrance. Yet we decided to take a broader and more generous view of secondary culture and to adhere to our principle of using the marks in all the subjects of the course.

We were encouraged to persist in this view by the reflection that all the subjects would have fairly high inter-correlations, and by certain results obtained by Mr Robert Wilkie, B.Sc. He prepared two lists of average first-year success marks for four of the senior secondary schools, the first being based only on the 'important' subjects (Latin, French, German, English, History, Geography, Mathematics, Science), the second being based on all the subjects. The average correlation between the two lists was +·985.

Importance of Comparability of Success Marks

This question of the comparability of the success marks is one of the most important in the whole technique of follow-up investigations. Its difficulty may be illustrated by considering the June Qualifying pupils who proceeded to the junior secondary courses. They distributed themselves among different schools, different courses in the same school, and different classes in the same course. We could not expect that the standards of the follow-up marks in the different schools or courses would be the same; therefore we had to deal separately with each course in each school. This, as we shall see, raises serious statistical problems if the attainment levels of the pupils in the different schools and courses vary. Furthermore, there is the possibility that the standards of marking will differ even in the various classes in the same course in the same school. These difficulties can, we believe, be overcome. But far the soundest policy is to evade them by taking every possible step to ensure that the follow-up marks are comparable in the first instance.

Steps taken to ensure Comparability of Success Marks

This question was discussed at various meetings with the head teachers of the senior and junior secondary schools and we concluded that there was no feasible method of ensuring uniformity in standard of follow-up marks for different schools, or for the different courses in the junior secondary schools. All we could reasonably attempt was to ensure that the follow-up marks would be comparable for the different classes in the same course in each junior secondary school, and that the marks would be comparable in the subjects common to the various courses in each senior secondary school.

This was done by an extension of a system already in general operation in many of the schools, whereby pupils in the different

classes of the same course all take the same terminal examination papers. To secure uniformity in the standard of marking many of the head teachers adopted the plan whereby each teacher, instead of correcting the papers of his own class, was assigned a certain number of questions in the papers of pupils in all the classes. From later inquiries Miss Young ascertained the exact procedure that had been used in each school and course, and it was then found that it had not always been possible to adopt this system. In certain schools, although uniform papers were set, each teacher corrected the scripts for his own class, the standards being then co-ordinated by the head of the department.

The final position, in short, was that in most schools we had full assurance that the follow-up marks were comparable; in others we had a fair measure of assurance; but in a number of schools there was no guarantee that the same standard had been used.

Check of Comparability of Follow-up Marks in the Various Classes

In the circumstances it was imperative that we should make a complete check to find which sets of marks should be excluded from our further calculations, and we did this in the following way. If, for the purposes of illustration, we consider the pupils who elected a particular course in one of the junior secondary schools, these would be arranged in, say, four classes, A, B, C and D. For these four classes we have the Qualifying examination means which are comparable. Now both Qualifying examination and follow-up marks are measures of scholastic attainment made at no great interval of time; and we may therefore assume that if the follow-up marks were comparable the follow-up means would be in the same order as the Qualifying means and separated by similar intervals.

If this assumption were sound, and if we were given the follow-up mean of Class A, we could calculate from the Qualifying data what the follow-up means of the other classes ought to be if the follow-up marks were comparable.[1]

[1] This is done by a method similar to that used when we transformed the teachers' estimates so that the mean and standard deviation of the estimates from a school were the same as those of the Qualifying marks of the pupils from that school. The formulæ are of the following form:

$$ {}_{B}M_{F} = {}_{A}M_{F} - \frac{\sigma_F}{\sigma_Q}({}_{A}M_{Q} - {}_{B}M_{Q}) $$

where ${}_{A}M_{F}$ = mean follow-up mark for Class A,
${}_{A}M_{Q}$ = mean Qualifying mark for Class A

The first purpose for which we used this method was to test the validity of the underlying assumption that the Qualifying means give a sound guide as to the relative positions of the follow-up means. This was possible because in many cases we knew that the follow-up means were comparable, and we could compare the calculated means with them. Data for a case of this kind are given below:

Class	F means (per cent.)		d	$pe_{\text{diff.}}$
	Actual	Calculated		
A	73·2	73·2	..	..
B	69·9	68·1	+1·8	3·3
C	68·5	67·6	+ ·9	2·9
D	66·7	64·7	+2·0	2·8
E	57·3	61·6	−4·3	3·5
F	49·6	51·3	−1·7	3·1

The figures reveal the width of the gap that separates the best from the poorest class, a point whose statistical import in follow-up work we shall discuss presently. They also show that the calculated means agree very closely with the actual means, the differences being in all cases explicable by the fallibility of the measures concerned.[1]

This method was applied to every class in every course in every school in all the three years of the follow-up, and the assumption stood the test in all cases where we had full assurance of the comparability of the follow-up marks. We therefore felt confident in using the check as a criterion for discarding groups for lack of comparability.

We give on p. 52 the data for a school in which we had not full assurance of comparability, but which we regarded as satisfying the test.

[1] The probable error which is relevant here does not concern the variations which arise through taking different samples of pupils from a total population; it relates to the fallibility of the *measurements*. Knowing the fallibility of the measures on which the calculated follow-up mean is based we can calculate its fallibility and the fallibility of the difference between it and the actual mean.

Reference: D. Brunt, *The Combination of Observations*. Cambridge: Cambridge University Press, 1917. P. 48.

The calculations are somewhat intricate and laborious, and in view of the number of cases with which we had to deal, and of the fact that great accuracy was neither possible nor necessary, the method used was to work out sample cases and interpolate by means of a graph.

Class	F means (per cent.) Actual	Calculated	d	pe_{diff}
A	64·6	64·6	..	..
B	61·3	57·6	+3·7	2·7
C	54·1	50·1	+4·0	2·6
D	50·9	49·4	+1·5	3·0

The following is a case where the test of comparability was not satisfied and the group of classes was therefore discarded from all further calculations:

Class	F means (per cent.) Actual	Calculated	d	$pe_{diff.}$
A	63·0	63·0	..	..
B	63·4	53·6	+ 9·8	2·4
C	58·3	45·4	+12·9	2·4

The amount of labour involved in this check was very great. We felt, however, that we were repaid by the confidence with which we could proceed with the further analysis of our data.

If we leave out the top classes, where the calculated mean was arbitrarily fixed to be the same as the actual mean, the result was that in 69 per cent. of the cases the differences between actual and calculated follow-up means were less than their probable errors; 96 per cent. of the differences were less than three times their probable errors. The number of rejections was therefore not very great.

A point which emerged from these calculations was that there is a general tendency for the follow-up means of the duller classes to be a little higher than one would expect on the basis of performance in the Qualifying examination. The difference is not large, the average value being 2 per cent., but it is almost universal. It may be due simply to greater leniency in marking the poorer classes, or to a tendency to push these classes nearer to the limit of their capacity.

Comparability of Follow-up Marks in the Various Schools

As has already been mentioned, we did not expect that the follow-up marks would be comparable between schools or between courses in the same junior secondary school. The latter was, indeed, no great disadvantage, for we wished to have separate

correlations for the various courses to see whether there were any significant differences. It may, however, be of some interest in connection with follow-up technique to show how the deviations in standard between school and school compare with those between class and class in the same school and course. We give below a comparison of actual and calculated follow-up means for the June group taking the boys' technical course in junior secondary schools:

School	F means (per cent.)		d	$pe_{\text{diff.}}$
	Actual	Calculated		
A	53·9	53·9	..	..
B	57·5	58·0	− ·5	1·4
C	58·5	53·8	+ 4·7	1·4
D	43·3	57·1	−13·8	1·4

Schools C and D differ in standard of marking by 18·5 per cent., and this holds almost exactly for all the courses in those two schools. In the commercial course, for instance, the difference is 18·4 per cent.

A comparison of actual and calculated follow-up means in senior secondary schools is shown below:

School	F means (per cent.)		d	$pe_{\text{diff.}}$
	Actual	Calculated		
A	64·8	64·8	..	..
B	61·7	69·3	−7·6	1·1
C	64·9	64·9	0·0	..
D	64·6	66·9	−2·3	1·3

The position is, therefore, that, thanks to the co-operation of the head teachers and staffs of the secondary schools, we could regard the follow-up marks of the classes in the same course in the same school as comparable, but that the standards as between school and school, and as between the different courses in the junior secondary schools, could not be regarded as comparable.

CHAPTER VII

METHODS USED IN COMBINING CORRELATIONS

THIS is a tiresome chapter, devoted to certain points of follow-up technique, which the general reader will be well advised to omit.

THE PROBLEM

Suppose we wish to find the best estimate of the correlation r_{QF}, and that we have six schools for which the follow-up (F) marks are not comparable though comparable within each school. Suppose also that the ability levels in the schools are different.

One method of finding an estimate of r_{QF} would be to calculate a correlation for the combined schools from a single grid, but this would involve an error caused by the differing standards of the follow-up marks. Another method would be to calculate separate correlations for the six schools and combine these by the usual z transformation. This, while avoiding the error of the first method, would introduce another, due to the differences in the ability levels of the schools. We shall call the first the error arising from differences in standard of marking, and the second, that arising from differences of the means.

The problem is to find a method of combining the correlations which avoids both sources of error.

SIZE OF ERROR CAUSED BY DIFFERENT STANDARDS OF MARKING

We can get a general idea of the error arising from differences in standards of marking as follows. If we have two large schools of the same size whose follow-up marks are on the same standard, the correlation for the combined schools would be

$$r_{QF}=\frac{\Sigma qf}{n\sigma_q\sigma_f}$$

If the follow-up marks for one of the schools were all raised by a per cent., the formula would become

$$\frac{\Sigma qf}{n\sigma_q\sqrt{\sigma_f^2+\frac{a^2}{4}}}$$

Thus a small variation in standard would not have a serious effect, but a difference of 15 per cent. might lower the correlation by as much as ·1.

The following correlations taken from Mr Bain's data give actual examples of the error in correlation caused by different standards of marking in one variate:

	English (N=858)	Arithmetic (N=853)	English and Arithmetic (N=842)
Slumped correlation with teachers' estimates	+·766	+·727	+·787
Slumped correlation with teachers' estimates scaled	+·879	+·799	+·883
Loss in correlation	·113	·072	·096

The differences in standards of marking are shown in Tables XIII and XIV, and we see that they give rise to a loss of approximately ·1 in correlation.

Size of Error caused by Differences in the Means

To illustrate the size of the error from the second source we take the six classes whose means are given on p. 51. In this case we know that the standards of marking are comparable, yet the difference of means is considerable, the average of the best class being 73·2 per cent. and that of the poorest 49·6 per cent.

A simple diagram will show that there would be a considerable error if we used the average of the correlations. A correlation can be represented pictorially by an ellipse which is a horizontal section of the normal bivariate surface. Omitting certain technical reservations about scales and standard deviations, we may interpret a broad ellipse as a low correlation and a long narrow ellipse as a high correlation. From the data for the six classes we can find the centres and the shapes of the ellipses and the positions of their major axes, and plot these on a diagram.

Fig. 6 shows the ellipses for the best and poorest classes. AB is the line of centres, and the other two lines are the respective

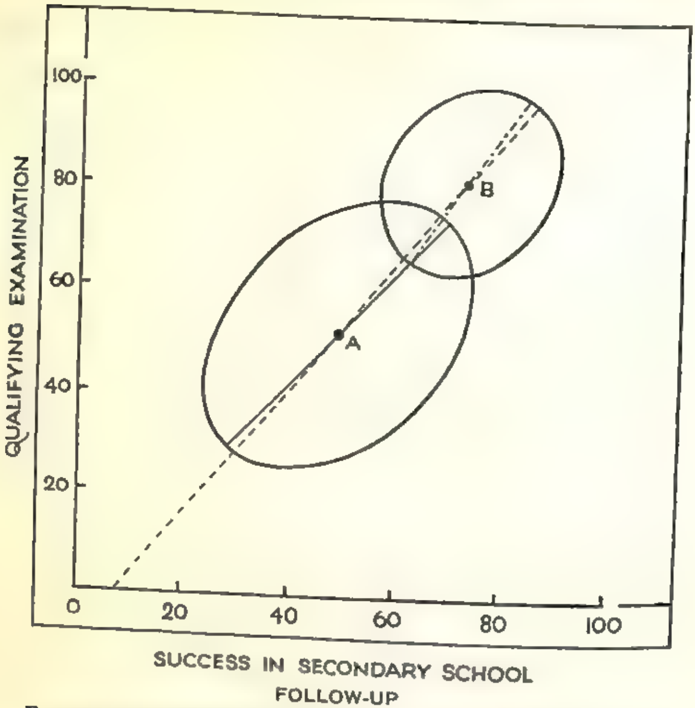

FIG. 6.—Effect of differences in the means when the differences are nearly the same in both variates.

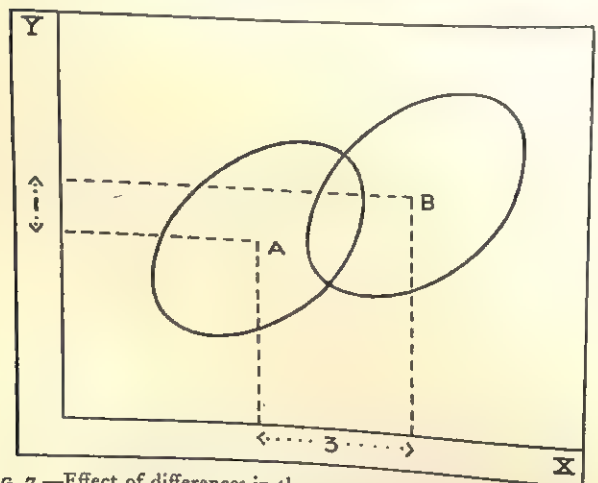

FIG. 7.—Effect of differences in the means when the difference in the y-variate is one-third of that in the x-variate.

major axes. The centres of the intermediate ellipses lie closely along AB.

While a single class gives a broad ellipse corresponding to a low correlation, the whole group would be represented by a much longer ellipse corresponding to a considerably higher correlation.

The correct correlation for the complete group, however, is not always higher than that for the homogeneous sub-groups; everything depends on the position of the means. If the difference in means of the y-variate were one-third of the difference for the x-variate, the position would then be that shown in fig. 7.

It is clear from this diagram that the correlations for the whole group may be the same as that for the sub-groups, or even lower.

To find the size of the loss in the case of these six classes, we calculated the correlation from a single grid and also by averaging the correlations for the separate classes. The results are as follows:

Average of correlations for six classes .	+·404
Correct correlation for whole group from single grid	+·619
Loss through averaging	·215

The loss in correlation in this way is well known,[1] but it is perhaps not generally realised that the difference could be as high as ·215

Methods of avoiding Error caused by Differences in the Means

The above examples show that the error *might* be serious; and the question now arises whether it would have to be taken into account, say, in combining correlations for the senior secondary schools, where the differences in means would be smaller. What we require is a method which will show whether a correction is required, and will enable us to make it if necessary.

USE OF A CORRECTION FOR RANGE [2]

It might be thought that we could overcome the difficulty simply by correcting the correlations for the separate groups for

[1] *Cf.* for example E. F. Lindquist, *A First Course in Statistics*. Boston: Houghton Mifflin Company, 1938. P. 179.

[2] *Cf.* G. H. Thomson, *The Factorial Analysis of Human Ability*. London: University of London Press, Ltd., 1939. Chapter xi.

T. L. Kelley, *Statistical Method*. New York: The Macmillan Company, 1923. P. 223.

range, but consideration of the possible arrangements of the ellipses will show that this would often be unsound. Further, if there were lack of comparability in both variates it would be impossible to obtain a satisfactory estimate of the standard deviation of the total population.

SCALING THE FOLLOW-UP MARKS

By using a process of scaling on Qualifying examination marks similar to that described in Chapter V, we can make the follow-up marks comparable and then calculate the correlation for the combined schools from a single grid. In Chapter VI we have given grounds for our belief that the underlying assumptions are legitimate, and another point in confirmation will be made at the end of the next section.

Instead of applying this method in the direct way, we found it better to use the following short-cut.

CALCULATION OF FOLLOW-UP CORRELATIONS WITH SCALED FOLLOW-UP MARKS WITHOUT ACTUAL SCALING

Since the scaling is done by equations, we can use these to find formulæ giving the values of the correlations with the scaled follow-up marks.

If $r_{IF(Q)}$ denotes the correlation between intelligence score and follow-up marks for all the schools, A, B, C, D, . . ., when the follow-up marks are scaled on Qualifying examination marks,

$$r_{IF(Q)} = \frac{\Sigma n_A \sigma_{IA} \sigma_{QA} r_{IFA} + \Sigma n_A ap}{\sqrt{\{\Sigma n_A \sigma_{IA}^2 + \Sigma n_A a^2\}} \sqrt{\{\Sigma n_A \sigma_{QA}^2 + \Sigma n_A p^2\}}}$$

where r_{IFA} = correlation between intelligence score and follow-up mark for School A,

n_A = frequency for School A,

σ_{QA} = standard deviation of Qualifying marks for pupils in School A,

a = deviation of mean intelligence score for School A from mean intelligence score for all schools,

p = deviation of mean Qualifying score for School A from mean Qualifying score for all schools,

Σ denotes summation for all schools, A, B, C, D, . . .

Formulæ [1] of this kind have some general interest. For instance,

[1] These formulæ can readily be put into terms of the correlations between the school means.

if we had calculated the correlations for the individual schools, we could find the true correlation for the combined schools. If the follow-up marks were known to be comparable, we could substitute σ_{FA} for σ_{QA}, and so on. The formulæ also show at a glance the effect of combining correlations when the means are the same but the standard deviations different. In this case, $a=b=\ldots=p=q=\ldots=0$.

As a test of the formulæ and of the assumptions on which the scaling system is based, we used one of them to calculate the true correlation for the six classes used as a previous illustration. From p. 57 it will be seen that the average correlation for the six classes is +·404 and that the true correlation is +·619. With the formula the result is +·618.

Application of Formula to determine Necessity for Correction for Difference in Means in Follow-up Correlations for Combined Schools

On applying the formula we found, as we expected, that the differences between the average and the true correlations were small, being only about ·007. Now, such a loss would be worth saving if we could rely upon our data to that degree of precision and if we were absolutely confident of the exactness of the method. Neither of these conditions for statistical tampering with results was, however, satisfied; and we decided, with relief, that it was sufficient to use simply the average values.

Application of Formula to determine Necessity for Correction for Difference in Means in Combining Correlations for June and December Groups

At no point in our investigation did we wish to be in the position of saying that our results were certain to be in error by some undetermined amount. We therefore applied the formulæ to another case where we knew there was a difference of means. This arose in the combination of the Qualifying data correlations for the December and June groups. If we call the correlations between English and Arithmetic for the separate groups ${}_{D}r_{EA}$ and ${}_{J}r_{EA}$, our problem was to find ${}_{(D+J)}r_{EA}$. In this case there was lack of comparability in both variates, and we knew that the June group was superior in quality.

TABLE XX

December and June Qualifying groups: Difference in intelligence and school attainment

	December group (N = 1,045)		June group (N = 1,940)		$M_J - M_D$
	M	σ	M	σ	
IQ	98·47	12·53	101·80	14·50	3·33
EQ *	98·22	13·65	101·67	14·10	3·45
AQ *	101·10	15·10	102·80	14·03	1·70

* EQ denotes English quotient and AQ Arithmetic quotient.

The differences in mean are statistically significant, and the question we have to answer is whether they are large enough to warrant the making of a correction in combining the correlations for the two groups.

The fact that both variates have to be scaled offers no mathematical difficulty, provided that we have a suitable basis for the scaling. In this case the only possible bridges are IQ, EQ and AQ, and we scaled I on IQ, Qe and E on EQ, and Qa and A on AQ.

If we call the correlation between English and Arithmetic, when English is scaled on EQ and Arithmetic on AQ, ${}_{(D+J)}r_{E(EQ)A(AQ)}$, the formula is found to be as follows:

$$ {}_{(D+J)}r_{E(EQ)A(AQ)} = \frac{\Sigma n_D \, {}_D\sigma_{AQ} \, {}_D\sigma_{EQ} \, {}_Dr_{EA} + \Sigma n_D ps}{\sqrt{\{\Sigma n_D \, {}_D\sigma^2_{EQ} + \Sigma n_D p^2\}}\sqrt{\{\Sigma n_D \, {}_D\sigma^2_{AQ} + \Sigma n_D s^2\}}} $$

where p and q = deviations of the English means of the December and June groups from the mean of the combined group,

s and t = deviations of the Arithmetic means of the December and June groups from the mean of the combined group.

Formulæ for other variates can easily be written down once the principle of their construction has been grasped; and the results of their application are given in Table XXI.

Again it appears that the error in our combined correlations arising from this cause is negligible.

TABLE XXI

December and June Qualifying groups: Correction of combined correlations for differences in means

Correlation	Correction
I and E	+·002
I and Qe	+·002
E and Qe	+·002
A and Qa	+·001
I and A	−·003
I and Qa	−·003
E and A	−·001
E and Qa	−·001
Qe and Qa	−·001
Qe and A	−·001

In more than half of the cases the correction is negative, showing that the method of averaging correlations sometimes gives an *over*-estimate even when both means for one group are higher than the corresponding means for the other. The explanation of this is given on pp. 56–57.

Variation of Correction with Size of Correlation and of Difference in Means

In the correlation r_{IQE} the difference of means in both variates was about 3·4 quotient points, and we concluded that a correction was not necessary. It is of some interest to find whether it would have been necessary if the difference had been, say, double the amount, and also how the correction would have been affected if the correlation had been lower. The size of correction for different sizes of correlation and of differences in means is shown below:

Size of correlation	Size of correction when difference in means is	
	3·4	6·8
+·844	+·002	+·008
+·600	+·004	+·020
+·300	+·009	+·036
·000	+·013	+·052

These results are not of general application, yet they show that the correction is larger for the smaller correlations and that its size rises steeply when the difference of means is increased.

Special Difficulty in Correcting for Differences in Means in Follow-up Investigations

In follow-up correlations we have seldom to deal with lack of comparability in the first variate. Our problem lies in the follow-up marks, and we have suggested that these should be scaled on the Qualifying examination. In doing so we are making the differences in mean the same for both variates in r_{QF}. It is, of course, very probable that they *ought* to be the same, but we are giving the benefit of the doubt to this one correlation, and thus putting it in a privileged position when we compare it with others.

It was partly for this reason that we decided to make no correction in our own follow-up, where all the corrections were small.

Conclusions in regard to Follow-up Technique

Our experience of this problem points to the following conclusions in regard to follow-up technique:

Every possible step should be taken to ensure that the follow-up marks of the classes are comparable within each school. If they are not comparable they may be scaled on Qualifying examination marks.

As far as the follow-up marks are concerned, it will normally be permissible to combine the correlations for the separate secondary schools by the z transformation, but before doing so the size of the correction for differences in means should be ascertained by the methods described above.

The experiment should be so planned that scaling of the first variate is seldom necessary; and such scaling should be done only on a bridge *in the same subject.*

Check that all Correlations which were combined could be regarded as derived from Equally Correlated Populations

This was done by the χ^2 method.[1] The total number of correlations involved was over 200, but it was usually obvious that the condition would be satisfied. The statistical check was

[1] *Cf.* P. R. Rider, *An Introduction to Modern Statistical Methods*, p. 106.

applied in doubtful cases, and only in four of these did P lie below ·01. Fortunately they fell in groups which were rejected on account of lack of comparability of follow-up marks, or concerned correlations with teachers' raw estimates.

Variance and Co-variance within and among Schools

The statistical reader will have noted the connection between the above formulæ and the methods of analysis of variance and co-variance, and it is possible that these methods could profitably be used in the study of the problems of this chapter. The variance for the five senior secondary schools is analysed below, the data used being the Qualifying examination marks:

	Degrees of freedom	Sum of squares	Mean square deviation
Within schools . .	364	70,466	194
Among schools . .	4	105	26
Total . .	368	70,571	

In this case $z = \frac{1}{2}\log_e w = 1{\cdot}00$.

The 5 per cent. value of z for $n_1 = 24$ is ·877, and the 1 per cent. value is 1·317. For $n_1 = \infty$, the 5 and 1 per cent. values are ·864 and 1·300. The obtained value for the significance of the variance among schools is therefore between the 5 and 1 per cent. points.

The largest variation among schools in all our follow-up data was that in the commercial course for the December group, and it was only after some hesitation that we decided to make no correction for differences of means even in this case. It was, however, a small group, and we felt that the absence of correction would make little difference in the final results. Below we give the analysis of variance and co-variance for this group of four junior secondary schools, the data being for the correlation r_{AF}:

	Degrees of freedom	Sum of squares A	Sum of products	Sum of squares F
Within schools .	120	13,277	5,824	10,847
Among schools .	3	1,674	712	804
Total .	123	14,951	6,536	11,651

This shows a much greater variation among schools, but even here the regression among schools is not significant.

CHAPTER VIII

COMBINATION OF CORRELATIONS

THE senior secondary group was distributed over five schools, while the main body of the junior secondary pupils went to four large schools, where they split up into three courses.

This complication of groups entailed the calculation of approximately 1,000 follow-up correlations, which have a somewhat oppressive effect if presented in the mass. We therefore give merely sample Tables to show the stages by which we arrive at the final combined results.

SENIOR SECONDARY SCHOOLS

ORIGINAL TABLES FOR SEPARATE SCHOOLS AND YEARS

TABLE XXII

Sample Table of correlations for a senior secondary school

Correlation with	$_1F_1$	$_1F_2$	$_1F_3$	$_1F_{Av}$	$_2F_{Av}$	$_3F_{Av}$
I	·715	·714	·680	·710	·653	·639
IQ	·696	·694	·703	·707	·599	·610
E	·580	·616	·586	·607	·581	·515
A	·700	·697	·698	·726	·675	·597
Qe	·743	·732	·731	·760	·636	·584
Qa	·662	·669	·695	·710	·665	·616
S	·707	·725	·709	·726	·694	·614
Q	·773	·771	·785	·809	·716	·660
T	·716	·685	·679	·688	·632	·572
Ts	..	..	..	·774	·654	·652

Even in this single school the correlations are reasonably steady, and many of the final results cast their shadows before. For

instance, Ts always shows a higher correlation than T; the correlations for Qe are always higher than those for E. This fact, that most of our results held not merely in the final combined Tables but for every school and for every year, added greatly to our confidence in them.

Only a very small number of the entrants to the senior secondary schools were from the December group, and these were omitted.

First Condensation: Combination of Schools

We had five Tables of the above type, and the next step was to reduce them to a single Table by combining the correlations for the five schools.

TABLE XXIII

Correlations for senior secondary schools combined

Correlation with	${}_1F_1$	${}_1F_2$	${}_1F_3$	${}_1F_{Av}$	${}_2F_{Av}$	${}_3F_{Av}$
I	·737	·701	·696	·710	·643	·646
IQ	·705	·671	·695	·690	·647	·640
E	·594	·576	·593	·595	·534	·529
A	·736	·697	·680	·690	·629	·567
Qe	·766	·733	·738	·749	·631	·593
Qa	·671	·672	·696	·686	·625	·597
S	·745	·711	·715	·717	·655	·615
Q	·802	·781	·800	·810	·714	·673
T	·705	·674	·680	·659	·616	·580
Ts	..	..	..	·737	·677	·640
Average r	·718	·691	·699	·704	·637	·608
N	367	359	367	368	333	259
σ_Q	..	..	..	13·83	12·97	11·15

Correction of Second- and Third-Year Correlations for Range

The differences between the correlations for the three terms of the first year are not significant statistically, so we may conclude that, on the average, the results do not establish a fall during the first year.

The slight rise from ${}_1F_2$ and ${}_1F_3$ to ${}_1F_{Av}$ might also be due to the fluctuations that are to be expected with relatively small numbers. There is, however, a mathematical explanation, in that the correlation with an average will usually be less than the average of the correlations with the elements which make up the average.[1]

We can also explain part of the fall from year to year. There is little or no leakage in the first year, but thereafter the numbers fall rapidly, due largely to the dropping out of weaker pupils. The narrowing of the range of talent, indicated by the standard deviations at the foot of Table XXIII, will cause a fall in the correlations.

We can correct for this narrowing of range,[2] and the correlations then approximate to what they would have been if *all* the pupils had gone on to the end of the third year. This has two advantages: it makes the correlations for the various years comparable, thus removing one obstacle to their further combination; and it enables us to show the true fall in predictive value from year to year. (See Table XXIV.)

Final Condensation: Combination of Correlations for the Three Years

The main aim of this chapter is to show the successive stages by which the correlations were condensed to single values for each measure, and further discussion of the fall from year to year is deferred until the corresponding results for the junior secondary schools are available. Until then, we shall also provisionally regard the other conditions for the combination of the correlations for the different years as satisfied. (See Table XXV.)

[1] The average of the correlations with I is

$$\frac{r_{IF_1} + r_{IF_2} + r_{IF_3}}{3}$$

The correlation with the average is

$$r_{I(F_1+F_2+F_3)}$$

which equals
$$\frac{r_{IF_1} + r_{IF_2} + r_{IF_3}}{\sqrt{\{3 + 2r_{F_1F_2} + 2r_{F_1F_3} + 2r_{F_2F_3}\}}}$$

if F_1, F_2 and F_3 are given equal weights.

Since $r_{F_1F_2}$, $r_{F_1F_3}$ and $r_{F_2F_3}$ will normally be less than 1, the second method would give a higher result than the first.

[2] *Cf.* p. 57. The narrowing of the range in follow-up marks was regarded as consequential upon the narrowing of the range in the first variate.

TABLE XXIV

Correlations for senior secondary schools combined: Second- and third-year results corrected for range

Correlation with	${}_1F_{Av}$	${}_2F_{Av}$	${}_3F_{Av}$
I	·710	·667	·724
IQ	·690	·671	·718
E	·595	·559	·611
A	·690	·653	·648
Qe	·749	·655	·674
Qa	·686	·649	·677
S	·717	·679	·695
Q	·810	·736	·748
T	·659	·640	·661
Ts	·737	·701	·718
Average r	·704	·661	·687

TABLE XXV

Final order of predictive value of measures for success in senior secondary schools

Order of merit	Final combined correlation with success
Qe	+·699
Qa	+·671
A	+·666
E	+·587
Q	+·770
Ts	+·720
I	+·700
S	+·698
IQ	+·691
T	+·653

Correlations for Batteries of the Original Measures

Few Education Committees would now, we hope, use a single test or examination for purposes of guidance or selection. They would be more likely to use a pair, like IQ + Q, or a battery of three, like IQ + Q + Ts. We have now, therefore, to find the order of merit for all the possible batteries,[1] and for the reasons offered in Chapter X we gave all measures equal weights.

The final correlations for batteries are given in Table XXXII, along with those for the junior secondary schools.

JUNIOR SECONDARY SCHOOLS

The results for the junior secondary schools are much more complicated and the number of original correlations much larger.

Original Tables for Separate Schools and Years

TABLE XXVI

Sample Table of correlations for a junior secondary school (Boys' technical course)

Correlation with	December group		June group	
	$_1F_{Av}$	$_2F_{Av}$	$_1F_{Av}$	$_2F_{Av}$
I	·606	·610	·670	·718
IQ	·569	·754	·650	·673
E	·419	·415	·536	·534
A	·627	·664	·548	·529
Qe	·632	·609	·709	·610
Qa	·548	·596	·545	·520
T	·641	·630	·517	·554
Ts	·708	·697	·806	·728

There were twenty such Tables.

[1] No references need be given on the methods used here; they will be found in any text-book of statistics.

The following is a sample formula:

$$r_{(I+Q+S+Ts)F} = \frac{r_{IF} + r_{QF} + r_{SF} + r_{TsF}}{\sqrt{\{4 + 2r_{IQ} + 2r_{IS} + 2r_{ITs} + 2r_{QS} + 2r_{QTs} + 2r_{STs}\}}}.$$

All the inter-correlations of the measures are therefore necessary, and these had to be obtained separately for the senior and junior secondary groups.

TABLE XXVII

Correlations for junior secondary schools combined

Correlation with	Boys' technical course				Girls' technical course				Commercial course					
	December		June		December		June		December			June		
	$_1F_{AV}$	$_2F_{AV}$	$_1F_{AV}$	$_2F_{AV}$	$_1F_{AV}$	$_2F_{AV}$	$_1F_{AV}$	$_2F_{AV}$	$_1F_{AV}$	$_2F_{AV}$	$_3F_{AV}$	$_1F_{AV}$	$_2F_{AV}$	$_3F_{AV}$
I	·604	·614	·621	·577	·645	·611	·619	·569	·553	·554	·714	·514	·504	·529
IQ	·576	·676	·605	·552	·635	·585	·590	·642	·536	·516	·773	·510	·495	·491
E	·470	·463	·549	·485	·548	·608	·479	·468	·446	·441	·596	·468	·421	·420
A	·478	·484	·463	·373	·557	·498	·584	·531	·475	·542	·752	·530	·505	·420
Qe	·590	·646	·654	·572	·588	·597	·614	·543	·572	·521	·635	·606	·568	·544
Qa	·485	·479	·510	·471	·563	·441	·506	·406	·398	·433	·468	·536	·486	·467
S	·564	·563	·660	·560	·656	·656	·615	·578	·582	·621	·852	·580	·538	·488
Q	·638	·667	·683	·612	·692	·624	·637	·540	·617	·606	·701	·660	·610	·585
T	·560	·609	·588	·586	·645	·570	·572	·506	·461	·549	·444	·568	·531	·598
Ts	·654	·686	·665	·629	·629	·638	·649	·596	·588	·659	·561	·668	·611	·608
Average r	·562	·589	·600	·542	·616	·583	·587	·538	·523	·544	·650	·564	·527	·515
N	276	174	492	359	231	114	310	219	125	94	27	217	200	119

We had four to six schools, three courses, and December and June groups. Only for the commercial course was it possible to obtain correlations for all three years of the follow-up; for the boys' and girls' technical courses the numbers in the third year were too small. (See Table XXVI, p. 68.)

First Condensation: Combination of Schools

The first step is to combine the schools after ascertaining that it is statistically legitimate to do so. When this is done we get the results shown in Table XXVII (p. 69).

Second Condensation: Combination of December and June Groups

The December and June groups can be regarded as equally correlated populations, and we therefore combine them, giving the results shown in Table XXVIII.

TABLE XXVIII

Correlations for junior secondary schools: Schools combined: December and June groups combined

Correlation with	Boys' technical course		Girls' technical course		Commercial course		
	$_1F_{Av}$	$_2F_{Av}$	$_1F_{Av}$	$_2F_{Av}$	$_1F_{Av}$	$_2F_{Av}$	$_3F_{Av}$
I	·616	·589	·631	·583	·529	·519	·568
IQ	·595	·595	·609	·624	·520	·502	·557
E	·523	·478	·509	·516	·461	·428	·457
A	·468	·410	·573	·521	·511	·517	·498
Qe	·632	·597	·604	·562	·595	·555	·561
Qa	·501	·475	·531	·418	·491	·410	·468
S	·628	·561	·633	·605	·581	·566	·579
Q	·667	·631	·661	·570	·645	·609	·606
T	·578	·593	·604	·528	·531	·536	·570
Ts	·661	·649	·640	·611	·640	·627	·599
Average r	·587	·558	·599	·554	·550	·527	·546
N	768	533	541	333	342	294	146

Correction of Second- and Third-Year Correlations for Range

The following correlations are those that would have been obtained if the whole of the first-year group had completed the course:

TABLE XXIX

Correlations for junior secondary schools: Schools combined: December and June groups combined: Second- and third-year correlations corrected for range

Correlation with	Boys' technical course		Girls' technical course		Commercial course		
	$_1F_{Av}$	$_2F_{Av}$	$_1F_{Av}$	$_2F_{Av}$	$_1F_{Av}$	$_2F_{Av}$	$_3F_{Av}$
I	·616	·621	·631	·606	·529	·515	·550
IQ	·595	·627	·609	·647	·520	·498	·538
E	·523	·509	·509	·539	·461	·423	·434
A	·468	·439	·573	·544	·511	·513	·477
Qe	·632	·629	·604	·585	·595	·551	·542
Qa	·501	·506	·531	·439	·491	·405	·445
S	·628	·593	·633	·628	·581	·563	·562
Q	·667	·662	·661	·593	·645	·607	·591
T	·578	·625	·604	·551	·531	·532	·552
Ts	·661	·680	·640	·634	·640	·625	·583
Average r	·587	·589	·599	·577	·550	·523	·527
Average r uncorrected for range	·587	·558	·599	·554	·550	·527	·546
N	768	533	541	333	342	294	146

Third Condensation: Combination of Correlations for Different Years

TABLE XXX

Correlations for junior secondary schools: Schools combined: December and June groups combined: Years combined

Correlation with	Boys' technical course	Girls' technical course	Commercial course
I	·618	·622	·528
IQ	·608	·624	·515
E	·517	·521	·442
A	·456	·562	·506
Qe	·631	·597	·569
Qa	·503	·497	·451
S	·614	·631	·571
Q	·665	·636	·621
T	·598	·585	·535
Ts	·669	·638	·624

Fourth Condensation: Combination of Correlations for Different Courses

TABLE XXXI

Final order of predictive value of measures for success in junior secondary schools

Order of merit	Final combined correlation with success
Qe	·605
A	·502
E	·499
Qa	·487
Ts	·648
Q	·645
S	·608
I	·597
IQ	·590
T	·578

We shall see in the next chapter that the statistical requirements for this combination are satisfied; but even if they were not, it would still have been desirable to make it, for only thus can we arrive at a final general order of merit of predictive value applicable to all post-primary courses. (See Table XXXI.)

Simplification of Tables for Batteries by Omission of Correlations with Intelligence Score and Teachers' Unscaled Estimates

Our next step is to find the correlations for batteries; but, in view of the results already given, there is no point in retaining data for the teachers' unscaled estimates, which should not be used for prognostic purposes.

After making this simplification we are still left with 23 different combinations of the fundamental measures, none of which includes both intelligence quotient and intelligence score. If we could justify the omission of either IQ or I we could reduce the number to 15.

COMPARATIVE PREDICTIVE VALUE OF INTELLIGENCE SCORE AND INTELLIGENCE QUOTIENT

The correlations for all the combinations, one of which includes intelligence score and the other intelligence quotient, were calculated, and the differences are shown below:

Battery	Correlation for battery with I, minus correlation for battery with IQ	
	Senior secondary schools	Junior secondary schools
I	+·009	+·007
I+Q . . .	+·014	+·002
I+S . . .	+·008	·000
I+Ts . . .	−·003	−·003
I+Q+S . .	+·009	·000
I+Q+Ts . .	+·004	−·001
I+S +Ts . .	+·001	−·003
I+Q+S+Ts . .	+·004	−·001

The average difference is ·004; none of the differences is significant statistically; I is higher in nine cases, IQ in five cases,

and in two cases the correlations are equal. Hence, so far as our data go, the two may be taken as equal in predictive value.

This point has a bearing upon the question of age-allowances, and also upon the difficulty created by the limited supply of tests with reliable norms. If we have only three or four from which to choose, we cannot avoid repetition and the possibility of candidates preparing for all of them. Since, however, intelligence score is as good as intelligence quotient, we may use an unstandardised test and thus widen our range of choice. Of course we could not then put the results in the familiar and easily interpreted form of IQ; but we feel that, in view of the lack of comparability between IQs determined by different group tests, this might be an advantage rather than a disadvantage.

With some hesitation we decided to omit intelligence score rather than intelligence quotient. This is not so sound statistically, for we mix scores and quotients; but the use of IQ is more in conformity with current practice.

Correlations for Batteries of the Original Measures

The correlations for the junior secondary schools are lower than those for the senior secondary schools. Now we should expect that the range of talent in the intellectual *élite* that enters the senior secondary schools would be narrower than that in the junior secondary courses; and that would no doubt be so to-day. In 1935–1936, with no strict selection, the standard deviation of the Qualifying examination marks for the senior secondary group was considerably larger than that for the junior secondary groups. We therefore applied a correction for range to the correlations for the latter, thus permitting the comparison shown in Table XXXII.

The agreement is remarkably close, both in size of correlation and in order of merit. It is, of course, not necessary that this should be so; and the differences are not necessarily due to sampling or experimental errors. It is quite possible that the order of predictive value may be different for the two types of school.

We started with nearly 1,000 correlations. These have now been condensed to 8, those for IQ, Q, S and Ts, and then expanded to 30. In performing this condensation we may, at times, have taken a little statistical licence, but Table XXXII is worthy of respect in consideration of the labour involved in its preparation.

TABLE XXXII

Predictive value of batteries for senior and junior secondary schools

Battery	Senior secondary schools	Junior secondary schools *	d
IQ+Q+Ts . .	·804	·806	−·002
IQ+Q+S+Ts .	·800	·799	+·001
Q +S +Ts . .	·790	·784	+·006
IQ+Q . . .	·786	·774	+·012
Q +Ts . .	·783	·780	+·003
IQ+Q+S . .	·783	·773	+·010
IQ+S +Ts . .	·782	·793	−·011
IQ+Ts . .	·779	·795	−·016
Q +S . . .	·774	·753	+·021
Q . . .	·770	·738	+·032
S +Ts. . .	·764	·775	−·011
IQ+S . . .	·736	·741	−·005
Ts . . .	·720	·741	−·021
S . . .	·698	·705	−·007
IQ . . .	·691	·689	+·002

* Corrected for range.

CHAPTER IX

PRELIMINARY DISCUSSION OF THE FOLLOW-UP CORRELATIONS

In this chapter we draw attention to certain points of general interest arising from the results just presented, and justify certain of the combinations we have made.

Is there a Fall in the Predictive Value of the Qualifying Measures in the Later Years of the Secondary Courses?

In the preceding chapter we have shown that there is a fall, but that this can be partly accounted for by the progressive narrowing of the ranges of the groups. We can now present in one Table all the data bearing on this question:

TABLE XXXIII

Average follow-up correlations for the first three years of the secondary courses *

Course	N	r		
	Year 1	1	2	3
Senior secondary . .	368	·704	·661	·687
Junior secondary:				
Boys' technical . .	768	·587	·589	..
Girls' technical . .	541	·599	·577	..
Commercial . .	342	·550	·523	·527

* The correlations for the second and third years have been corrected for range.

When we apply the statistical tests [1] we find that none of the differences is statistically significant.

[1] For the senior secondary and commercial courses the χ^2 test is the most convenient. For the senior secondary courses, where the differences are greatest, χ^2 is nearly 1·2: P is therefore about ·5. For the girls' technical course the z difference is ·034: SE=·019.

While the differences in average correlations are not significant, it is possible that those for the individual measures might be. This was tested for the four main measures, IQ, Q, S and Ts, and it was found that, on the whole, the correlations might be reasonably regarded as coming from equally correlated populations.[1]

These results give some statistical justification for the combination of the correlations for the different years; and, while they do not rule out the possibility of a real fall from the first to the third year, they show that it is unlikely to be much in excess of ·02. The point is of great importance, for if the fall had been considerable, we should have had to recast all our ideas as to the possibility of selection at the Qualifying stage. Measures which give a good prediction of success in the first year only would be of little value.

Rise in Predictive Value of IQ in Later Years of Courses

An examination of Tables XXIV and XXIX gives little or no evidence of change in the relative position of I and IQ in the later years of the courses; but it does confirm the view that intelligence will tell in the end, and that the IQ rises in predictive value relative to the other measures as the years go on. In every course the correlation of IQ with F rises from the first to the last year, while the correlations for S and Q always fall. The differences are not great, but their consistency adds to our confidence in their existence. A comparison of correlations for the first and last years for all secondary courses is given below:

Measure	Average rise or fall from first to last year
IQ	Rise ·029
Ts	Fall ·016
S	Fall ·020
Q	Fall ·047

The puzzling fact that the correlations are higher for the third than for the second year may be due to chance fluctuations or to the correction for range. There may, however, be other reasons. The pupils may work more consistently up to their capacity, and

[1] The most doubtful case in the group for the senior secondary schools is that of Q, where $\chi^2 = 7{\cdot}3$, and P lies between ·02 and ·05. In the junior secondary group the only difference which approaches significance is that for Q in the girls' technical course. Here the z difference is ·112: SE = ·061.

there may be increased steadiness in the teachers' standards of marking as the Leaving examination approaches. Again, the pupils who have left comprise for the most part those not interested in the secondary course, and in this group the correlation between Qualifying attainments and success would probably be low.

This tendency for the third-year correlations to be higher than those for the second makes it difficult to give an exact statement as to the significance of the rise in the follow-up correlations with IQ relative to those for the other measures, as in some cases the second is the last year, and in others the third is the last. If we consider Q, the fall from the first to the third year in the senior secondary group is just significant;[1] and while the rise for IQ is not in itself significant, the difference in relative position of IQ and Q in the last year is decidedly so. We may therefore take it as very probable that in the later years of the course IQ rises in predictive value relative to the other Qualifying measures.

Does the Order of Predictive Value change in Later Years of the Secondary Course?

TABLE XXXIV

Changes in order of predictive value of the measures in the first three years

Measure	Place in order of predictive value									
	Senior secondary			Junior secondary						
				Boys' technical course		Girls' technical course		Commercial course		
	Year 1	2	3	1	2	1	2	1	2	3
Q	1	1	1	1	2	1	4	1	2	1
Ts	2	2	2	2	1	2	2	2	1	2
S	3	3	4	3	4	3	3	3	3	3
IQ	4	4	3	4	3	4	1	4	4	4

[1] *Cf.* footnote, p. 77.

On the whole the measures retain their places with reasonable consistency, especially when it is remembered that a change in order may be due to a very small change in the relative sizes of the correlations which determine it. As we should expect, IQ shows a faint tendency to improve its place in the later years.

The answer to the question therefore appears to be that there is no clear change in the order, but that there is a change in the spacing.

Are the Correlations Higher for some Courses than for Others?

From the figures given in Table XXXIII it would seem that the correlations are highest for the senior secondary courses, very nearly equal for the boys' and girls' technical courses, and lowest for the commercial course. But when we take account of the relative ranges, we find that there is no evidence of any difference between the senior secondary, the boys' technical and the girls' technical courses. There is, however, a statistically significant difference between these and the commercial course,[1] and once again our confidence in this curious result is increased by the fact that it appears not only in the averages, but separately for the December and June groups, and also for each separate year.

Does the Order of Predictive Value vary from Course to Course?

TABLE XXXV

Order of predictive value of the measures for the different courses

Measure	Senior secondary	Junior secondary		
		Boys' technical	Girls' technical	Commercial
Ts	2	1	1	1
Q	1	2	2	2
S	3	3	3	3
IQ	4	4	4	4

[1] If we take the first-year correlations for the girls' technical and the commercial courses the z difference of the average correlations is ·073: SE = ·023.

Our answer is, therefore, that the predictive value of the Qualifying measures is the same for the senior secondary, the boys' and girls' technical courses, but slightly lower for the commercial course.

The order is even more consistent than for the different years of the same course, and the only notable point is that Q takes first place in the senior secondary course, while Ts gives the highest correlation for all the junior secondary courses. If we take the separate senior secondary schools and years we have fifteen comparisons between Ts and Q, and in all but three of these Q takes first place; in all but four of fourteen comparisons for the junior secondary schools Ts takes first place.

The answer to our question is, therefore, that there is a probability that Q is better for predicting success in the senior secondary schools than Ts, but that otherwise the order of the measures is the same in all types of secondary course.

CHAPTER X

THE WEIGHTING OF MEASURES

Nominal and Actual Weights

We first warn the general reader against two sources of misunderstanding:

> If Qe, with a possible mark of 200, and Qa, with a possible mark of 100, are combined to give Q; or
>
> If the possible marks in Qe and Qa are 100, and if, before combining, we double the English marks;

in neither case are we necessarily giving English twice the weight of Arithmetic.

The reason is, that the actual weights depend upon the relative dispersions of the marks combined. If σ_E and σ_A are the standard deviations of the English and Arithmetic marks, and if, before combining, we multiply the marks by the *nominal* weights, W_E and W_A (giving the combined score $W_E E + W_A A$), then the *actual* weights given are $W_E \sigma_E$ and $W_A \sigma_A$.

When we combined Qe and Qa to give Q, the weighting was not $1\frac{1}{2} : 1$, but a complex one depending upon the standard deviations, and specific to our own examinations. It was for this reason that, in finding correlations for the batteries, we used the simple standard system of equal weights.

We must now consider the questions, whether a better prediction would be obtained by using a different weighting system, and whether there are any general rules in regard to weighting which Education Committees should follow. For this purpose we use a statistical technique [1] which enables us to find the weights giving the highest correlation with F, called the *best weighting*, and the size of the correlation for the best weighted battery.

Weighting of English and Arithmetic

The problem is to find what weights should be given to English and Arithmetic in combining Qe and Qa to give Q, or in combining

[1] G. H. Thomson, "On the Computation of Regression Equations, Partial Correlations, etc.," *British Journal of Psychology*, vol. xxiii, Part I, July 1932, p. 64.

E and A to give S. To solve it Dr McIntosh and Miss Young calculated the best weighted correlations for the combined senior secondary schools and found the differences between these and the equal weighted correlations. The improvement effected is shown below:

Combination	Difference between best weighted and equal weighted correlations with F					Best weighting				
	${}_1F_1$	${}_1F_2$	${}_1F_3$	${}_1F_{Av}$	${}_2F_{Av}$	${}_1F_1$	${}_1F_2$	${}_1F_3$	${}_1F_{Av}$	${}_2F_{Av}$
Qe+Qa	·005	·005	·001	·002	·003	3 : 2	3 : 2	1 : 1	3 : 2	2 : 3
E +A	·007	·006	·007	·007	·009	2 : 3	2 : 3	2 : 3	2 : 3	1 : 2

The differences are, of course, all in favour of the best weighted correlations, and they show that the improvement resulting from best weighting is very small. When we take this in conjunction with the facts that the best weightings are different for the tests and for the examinations, that they vary even for the same pair of examinations, and that the best weightings given above are specific to our particular tests and examinations and would not necessarily hold for others, we may conclude that the weighting of English and Arithmetic in Qualifying tests or examinations need not be attempted.

Batteries—Best Weighting

The improvement effected by the best weighting of batteries, as shown by the increase in the follow-up correlations, is as follows:

Battery	Senior secondary schools		Junior secondary schools [1]	
	Improvement	Weighting	Improvement	Weighting
I +Q+Ts	·004	5 : 7 : 3	·002	6 : 5 : 8
I +Q+S+Ts	·008	6 : 9 : 1 : 4	·010	9 : 1 : 8 : 3
I +Q	·005	4 : 7	·003	2 : 3
I +Q+S	·013	6 : 11 : 1	·007	8 : 3 : 12
Q+S +Ts	·004	7 : 3 : 4	·003	3 : 4 : 6
I +S +Ts	·003	5 : 4 : 7	·009	11 : 5 : 16
Q+Ts	·004	4 : 3	·000	1 : 1
I +Ts	·000	5 : 6	·002	5 : 7
Q+S	·008	3 : 2	·003	1 : 2
S +Ts	·001	9 : 11	·002	2 : 3
I +S	·000	1 : 1	·000	4 : 5

[1] The calculations for the junior secondary schools were made by Miss M. Mitchell, M.A.

The improvement is very small. It would be well worth making if we could assign systems of best weights for various batteries which would be generally applicable, but the above results show that this is impossible; even with the same tests there are striking variations between the senior and junior secondary groups.

Mr A. S. Robertson, M.A., using a group of four senior secondary schools, found the following variations in best weighting for the same tests and the same group of pupils:

Battery	Best weightings for correlation with				
	$_1F_1$	$_1F_2$	$_1F_3$	$_1F_{Av}$	$_2F_{Av}$
IQ + Qe + Qa .	5 : 12 : 7	3 : 13 : 11	4 : 11 : 9	4 : 12 : 9	6 : 9 : 13
IQ + E + A .	9 : 4 : 9	6 : 5 : 10	8 : 4 : 9	8 : 4 : 10	4 : 5 : 11
Qe + Qa + T .	13 : 6 : 5	13 : 7 : 3	11 : 8 : 5	12 : 7 : 5	9 : 12 : 2
E + A + T .	9 : 10 : 14	9 : 11 : 12	9 : 10 : 14	9 : 10 : 15	8 : 11 : 11
IQ + Qe + Qa + T	2 : 4 : 2 : 2	1 : 5 : 3 : 1	1 : 4 : 3 : 2	1 : 4 : 3 : 2	1 : 3 : 5 : 8
IQ + E + A + T .	3 : 2 : 3 : 4	2 : 2 : 3 : 4	3 : 2 : 3 : 4	2 : 2 : 3 : 5	1 : 2 : 4 : 4

Some of Mr Kidd's results show that the best weightings for a given battery vary considerably from school to school:

School	Best weighting for correlation of battery I + Q with			
	$_1F_1$	$_1F_2$	$_1F_3$	$_1F_{Av}$
A	2 : 3	1 : 2	1 : 4	1 : 2
B	1 : 3	2 : 5	2 : 5	1 : 8
C	2 : 5	2 : 3	3 : 4	2 : 5

In view of these results we find it impossible to lay down a single principle even of the most general character. Since, in addition, the gain by any system of best weighting is very small, we recommend that Education Committees should simply take an ordinary average of the results of the tests which they employ. The time and labour involved in the calculation of averages with complicated weights would be better devoted to the scaling of the teachers' marks.

CHAPTER XI

CORRELATIONS WITH THE BEST BATTERY

WHEN no follow-up data are available as a criterion, a rough idea of the relative values of the various measures can be obtained by assuming that the battery of all the tests is likely to be the best, and calculating correlations with it.

The method has too much of the character of a vicious circle to command great confidence, but since we have a follow-up

TABLE XXXVI

Correlations of batteries with best battery IQ + Q + Ts

Battery (1)	Correlation with best battery IQ + Q + Ts (2)	Correlation with best battery (on scale of follow-up correlations) (3)	Correlation with F (4)	*d* (5)
IQ + Q + S + Ts	·996	·799	·799	·000
Q + S + Ts .	·987	·787	·787	·000
IQ + S + Ts .	·987	·787	·787	·000
IQ + Ts . .	·992	·794	·787	+ ·007
Q + Ts . .	·980	·777	·781	− ·004
IQ + Q . .	·985	·784	·780	+ ·004
IQ + Q + S .	·984	·782	·778	+ ·004
S + Ts . .	·974	·768	·769	− ·001
Q + S . .	·979	·775	·763	+ ·012
Q . .	·970	·762	·754	+ ·008
IQ + S . .	·955	·740	·738	+ ·002
Ts . .	·944	·724	·730	− ·006
S . .	·938	·715	·701	+ ·014
IQ . .	·921	·690	·690	·000

criterion our data allow us to make a test of its validity. For this purpose we used the best battery, IQ + Q + Ts, rather than the complete battery.[1]

Our interest lies in the way in which the order so obtained compares with that of the follow-up correlations. The latter, which are given in column (4) of Table XXXVI, are the average correlations for senior and junior secondary schools; and to make the comparison easier we have reduced those with the best battery to the same scale. This was done roughly by equating the highest and lowest values and placing the intermediate points by means of a straight-line graph.

The conformity between the two orders is remarkable; it is, in fact, much closer than that between the follow-up correlations for the senior and junior secondary schools. The average difference, without regard to sign, is ·004, as compared with ·011 for the differences in Table XXXII.

These results give some justification for the belief that the method has evidential value; and, in view of the labour involved in large follow-up experiments and the time that must elapse before results are available, all ways of making legitimate short-cuts are important. Methods of this kind have, of course, serious limitations. If a new battery were prepared we could find its correlation with IQ + Q + Ts. If this came out to be ·996 we could say that it was probably a very good battery, but we could not say whether it would give a higher or lower correlation with success. If, on the other hand, the correlation were ·4, it would probably be a poor battery.

[1] The correlations were not calculated directly from grids but by means of formulæ of the type mentioned in the footnote on p. 68.

CHAPTER XII

PROFILES: A SECOND METHOD OF DETERMINING THE PREDICTIVE VALUE OF MEASURES

EACH child at the Qualifying stage sets a problem: by the end of a follow-up inquiry we know the correct answer. We can therefore study any suggested procedure of selection, using as our test the number of cases in which our answers would have been wrong.

LIMITATIONS OF THE CORRELATION TECHNIQUE

The correlation technique, which we have used so far, has many advantages, but it involves an immense amount of labour and has serious limitations. On the basis of the correlation of the various batteries with success we should expect to make the fewest mistakes in selection by using IQ + Q + Ts, but this is not absolutely certain. The correlations may not be uniform throughout the range: a battery may be highly selective at the top of the scale and less selective at the border-line. Further, the evaluation of a battery by correlation ignores the possibility of improving its selectiveness by considering, in border-line cases, the way in which the total score of a pupil is made up.

Another serious limitation of the correlation technique is that, by itself, it gives us no information on the very important practical problem of fixing the pass-mark, a point which we shall postpone to a later chapter.

For these and other reasons given later we decided to work over our data again, using an entirely different technique which throws light on the blind spots of the selection problem left by the correlation procedure.

PROFILES

For this alternative method it was necessary to have the detailed Qualifying and follow-up data for each pupil in a form in which any required fact could be seen at a glance. We therefore decided to make graphical profiles of the type shown in fig. 8 (p. 93) for each of the senior secondary pupils. As the technique is one which

we recommend to future investigators we shall deal with their construction in some detail.

At the left-hand side we have the pupil's scores in the seven Qualifying tests and examinations. The next point, marked Av, gives the unweighted average of the seven. Then follow three points giving the primary teacher's estimates of the pupil's practical ability (Tp), industry (Ind.), and the Medical Officer's grading for health (H).

On the right-hand side, the point between the first pair of vertical lines is the primary teacher's forecast of the pupil's success *in the post-primary course actually selected.* The next three points, marked I, II and III, give the pupil's performance in the first three years of the post-primary course; and the last point indicates the head master's final estimate of success.

Necessity for using Standard Scores

Ordinary marks or actual scores would be unsuitable for the construction of such profiles. Even if all scores were in percentages they would not be comparable and would be misleading. If, for example, a pupil had 70 per cent. in English and 70 per cent. in Arithmetic, the graph would give the impression that he was equally good in both; but this would clearly not be the case if the mean score in English were 70 per cent. and that in Arithmetic 60 per cent. If, therefore, we wish the scores to be comparable, we must relate them to their respective means. Such a procedure will enable us to use a fixed line of means running right across the graph, so that we can tell immediately whether the pupil is above or below average in any particular measure. Even this method would be misleading if there were differences in the standard deviations. If a pupil is 10 per cent. above the average in English and in Arithmetic it does not follow that he is equally good in both; this depends on the standard deviations of the marks in the two subjects.

For true comparability, therefore, we must take account both of the means and of the standard deviations, and this is done by the use of what are called standard scores. These are obtained by finding the pupil's deviation from the average score of his group and dividing it by the standard deviation of the group. If the mean and standard deviation of the group are 60 per cent. and 10 per cent. respectively, and if a pupil's score is 70 per cent., his standard score is thus $+1\sigma$.

The use of standard scores results in profiles in which all scores and all differences are comparable, and there is the further advantage that every point has a definite significance in relation to a normal distribution. If a pupil is placed at $+1\sigma$, this means that, in a normally distributed group, only about 16 per cent. of the pupils would have a higher score; if at -1σ, only about 16 per cent. of the pupils would have a lower score. If his placing is $+2\sigma$, only about 2·3 per cent. of the pupils would have a higher score; and a standard score of $+3\sigma$ may be taken to indicate that the pupil is the best in his group.

This, then, is the method which we adopted in making the profiles. The horizontal at the middle of the graph is the average line, and the dotted horizontals are drawn at intervals of 1σ. In fig. 8 we see at once that the pupil is of outstanding intelligence. He is, relatively, not so good in English; and we have a confidence in the reality and size of the difference that we could not have had if we had used ordinary marks.

Calculation of Standard Scores for Success Marks

There are many technical difficulties in the determination of standard scores for our present purpose, which can best be illustrated by considering the ${}_3F_{Av}$ marks. We wish to deal with the group as a whole, and the scores must therefore be strictly comparable for all pupils.

The pupils are in five different senior secondary schools whose follow-up marks are not comparable. If we simply calculate a pupil's standard score from the mean and standard deviation *of his own school*, there is an error through assuming that the average standards of intelligence and attainment in all schools are the same. The standard ${}_3F_{Av}$ scores have therefore to be corrected for differences in the average levels of the schools by arranging that they are given relatively to the mean and standard deviation of the complete group, that is, the pupils in all the schools if they were all marked on the same standard. All the ${}_3F_{Av}$ marks are then comparable among themselves.

We also wish the ${}_3F_{Av}$ marks to be comparable with those for ${}_1F_{Av}$ and ${}_2F_{Av}$; but the third-year group is more highly selected than the groups for the first two years on account of the dropping out of weaker pupils. If, therefore, a pupil has a standard score of $+1\sigma$ in the third year he is better than one who has a score of $+1\sigma$ in the first year. This difficulty can be overcome by making

the scores for all years relative to the first-year group. That is to say, when we give a transformed score of $+1\sigma$ for a third-year pupil this means that, if the whole first-year group had remained to the third year, he would have been $+1\sigma$ above the average of this group. Otherwise a pupil might appear to be falling back in his secondary work, whereas he is actually improving.

Starting, then, from the raw ${}_3F_{Av}$ marks we have to make three transformations: (1) into standard scores, (2) to ensure comparability between school and school, and (3) to ensure comparability between year and year.

The foundation of the second and third corrections is the scaling of the follow-up marks of the different schools on the comparable Qualifying examination marks. That is to say, the transformed follow-up marks for a school will have the same mean and standard deviation as the Qualifying examination marks for the pupils of that school. We can now treat the follow-up marks for all the schools as comparable, and use the mean and standard deviation of the whole group for fixing the standard scores. It is, however, not necessary to scale the marks as a separate step. Each of the three transformations is linear; therefore we can make all of them at once by means of a single equation which, though more complicated, is still linear.[1]

[1] Starting from the fundamental scaling equation (p. 44), and knowing the nature of the transformations to be made, there is no difficulty in arriving at the following formula for the complete transformation of a raw third-year score ${}_AF_3$ for a pupil in School A into a comparable standard score ${}_AX_3$:

$${}_AX_3 = \frac{M_{Q_3} - M_{Q_1}}{\sigma_{Q_1}} + \frac{a_3}{\sigma_{Q_1}} - \frac{{}_A\sigma_{Q_3}\,{}_AM_{F_3}}{\sigma_{Q_1}\,{}_A\sigma_{F_3}} + \frac{{}_A\sigma_{Q_3}}{\sigma_{Q_1}\,{}_A\sigma_{F_3}} \cdot {}_AF_3$$

where M_{Q_3} = mean Qualifying examination score of the whole ${}_3F_{Av}$ group for the five schools,

M_{Q_1} = mean Qualifying examination score of the whole ${}_1F_{Av}$ group for the five schools,

a_3 = deviation of mean Qualifying examination mark of ${}_3F_{Av}$ group in School A from the mean Qualifying examination mark for the ${}_3F_{Av}$ group in all five schools,

${}_A\sigma_{F_3}$ = standard deviation of ${}_3F_{Av}$ marks for School A.

The equation is linear, and when we put in the values of the constants it reduces to

$${}_AX_3 = -3{\cdot}631 + {\cdot}068\,{}_AF_3$$

Thus, if a pupil's raw ${}_3F_{Av}$ mark is 80 per cent., his comparable score is $+1{\cdot}8\sigma$.

Different equations, simpler in form, are required for the first and second years; 15 equations in all are required.

Calculation of Standard Scores for Qualifying Tests and Examinations

The first seven data on the left-hand side of the profiles are based on uniform tests or examinations, and are comparable within the December group and within the June group. We knew, however, that the level of intelligence and attainment was higher for the June than for the December group, a difficulty which we surmounted by the method of scaling, using $P = \frac{IQ + EQ + AQ}{3}$ as a bridge.[1]

Placing the Five-Point Gradings on the Profiles

It was not necessary to place the gradings for industry, health, etc., with great accuracy, but it would have been misleading if we had placed, say, all the A gradings at the same σ distance above the line of means. If, for example, 1 per cent. of the pupils were graded A in health and 10 per cent. in industry, an A in health would indicate a higher standard than an A in industry, and the A point for health should be higher above the line of means. By a statistical procedure we can, from the distribution of the gradings, find the means and the ends of the ranges for the various categories, and these are indicated in the profiles by points with vertical lines through them.[2]

[1] The following typical equation is that for converting Qe marks for the December group into comparable standard scores:

$$X = \frac{a}{\sigma_P} - \frac{{}_D\sigma_P \, {}_DM_{Qe}}{\sigma_P \, {}_D\sigma_{Qe}} + \frac{{}_D\sigma_P}{\sigma_P \, {}_D\sigma_{Qe}} \cdot Qe$$

where Qe = raw mark,
X = standard score,
${}_DM_{Qe}$ = mean Qe mark for the December group,
$a = {}_DM_P - {}_{(D+J)}M_P$.

[2] In these calculations we assume that the measures are normally distributed. We find the feet of the ordinates which would divide a normal distribution in such a way that the areas of the sections would correspond to the percentages of pupils graded in the various categories. Finally we calculate the means of the sections of the curves, which will, of course, not usually be at the mid-points of the ranges. For the primary teacher's forecast and the final success grading, where the letters had a definite objective meaning, we did not use the actual percentages in the various categories; we simply divided the base-line of a normal distribution into five equal parts.

References: K. J. Holzinger, *Statistical Methods for Students in Education*, p. 219.

S. Dawson, *An Introduction to the Computation of Statistics*. London: University of London Press, Ltd., 1933. P. 93.

Calculation of the Average of the Seven Qualifying Tests and Examinations

The average of the seven Qualifying tests and examinations is found by first adding the corresponding scores and then dividing by 7. Since each element in the numerator is in standard scores it might be thought that the averages obtained in this way would also be in standard scores; but this is not so. Their mean would certainly be zero, but the standard deviation would be less than unity by an amount which depends on the size of the inter-correlations. We had, therefore, to transform the averages into true standard scores by dividing by their standard deviation.

STANDARD SCORES FOR QUALIFYING DATA NOT COMPARABLE WITH THOSE FOR SUCCESS

There is one sense in which the data on the profiles are not comparable. A pupil may be above $+1\sigma$ in the Qualifying data, but well below the average line in the success data. At first sight it would appear that he has fallen off very badly. It must be remembered, however, that his scores on the left-hand side show that he is above the average *of the whole Qualifying group*; while on the right-hand side his scores are below the average *of the highly selected group that enters the senior secondary courses.*

To avoid confusion we shall now denote the standard deviation of Qualifying scores for the complete Qualifying group by sigma, using σ for the standard deviation of the post-primary group with which we are dealing.

The Short Horizontal Lines on the Profiles

There are two short horizontal lines on the profiles. The one at the left-hand side is drawn horizontally at $+\cdot 7$ sigma, and may be regarded as an entrance pass-mark for the senior secondary courses. Pupils whose average scores lie above it have more than a 50 per cent. chance of success. Its determination will be explained later. If a pupil's success scores lie above the short horizontal at the right, it may be taken that he has made good in the secondary course.

Sample Profiles

As these profiles have more human interest than the Tables of correlations with which we have hitherto wearied the non-mathematical reader, we shall now give a few samples which will afford

practice in their interpretation. We shall deal first with standard cases, where the pupil's performance in the secondary course is in conformity with his Qualifying data.

THE PUPIL WITH THE HIGHEST QUALIFYING SCORE

Fig. 8 is the profile of a pupil who has fully borne out his early promise; he set us an easy problem at the Qualifying stage, and our answer would have been correct. This is a boy on whom fortune has smiled. He comes from a well-to-do home, has an imposing list of good qualities, 'perseverance, originality, initiative and self-reliance,' and his primary teacher rightly described him as a boy of 'great possibilities.' The fact that Music, Art and Bible are the only subjects of the secondary course which he regards as difficult is, we suspect, an indication of his tastes rather than of his abilities. He obtained his Leaving Certificate with great credit in 1941.

THE PUPIL WITH THE LOWEST QUALIFYING SCORE

For comparison we give, in fig. 9, the profile of the pupil who had the lowest average score in the whole Qualifying group. His follow-up scores are relative to the average of the backward class to which he was sent. The two form a striking contrast and should be of interest to any who still believe that all children are born equal in capacity, or that they should all have the same post-primary education.

A CLEAR ADMIT

The profile in fig. 10 is that of a pupil whom we should admit without hesitation. His standard follow-up scores of $+\cdot4\sigma$ to $+\cdot5\sigma$ are just what one would expect from his Qualifying tests.

A BORDER-LINE ADMIT

The next profile (fig. 11) is that of a girl of good ability in English but weak in Arithmetic. She is just above the border-line, and our hesitation as to her fitness for admission is increased by the low health grading. We find her, as we should expect, just above the success line in the secondary course.

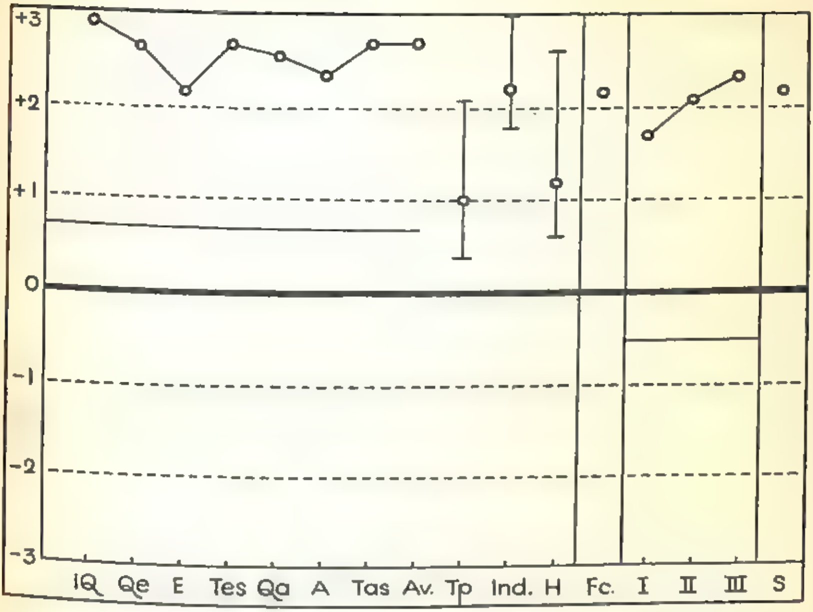

Fig. 8.—The pupil with the highest Qualifying score.

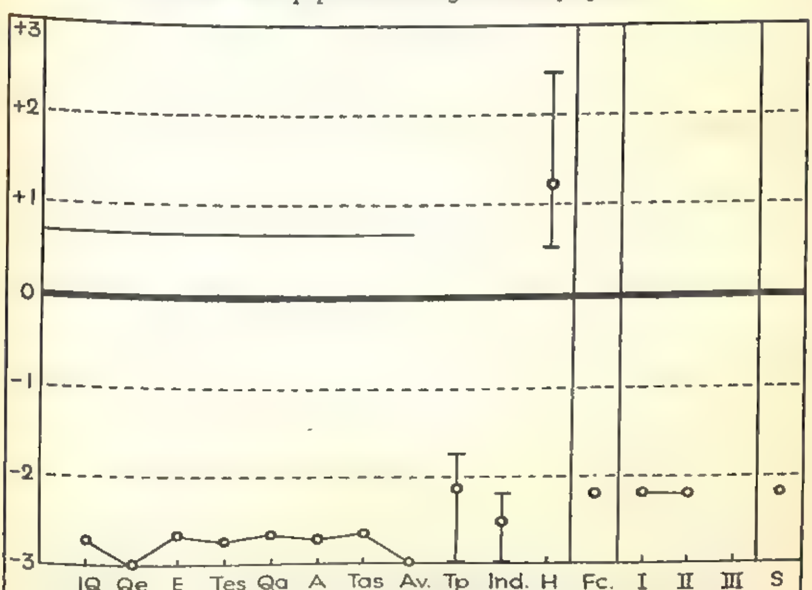

Fig. 9.—The pupil with the lowest Qualifying score.

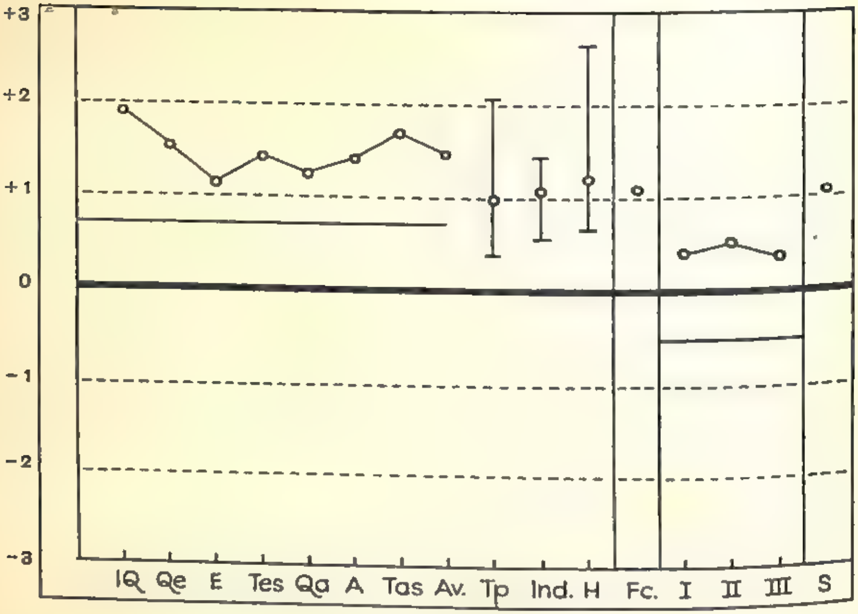

FIG. 10.—A clear admit.

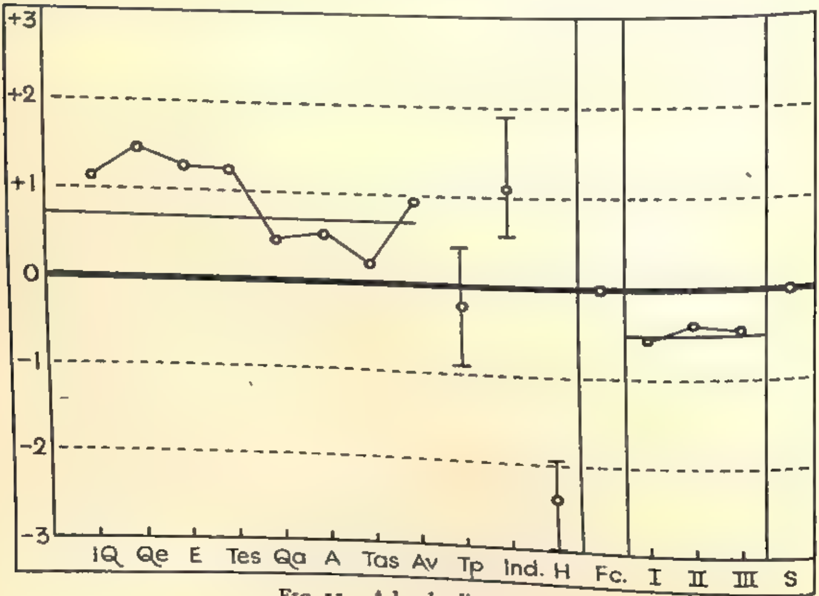

FIG. 11.—A border-line admit.

A CLEAR REJECT

The next (fig. 12) is a case where we should reject without hesitation. This girl is clearly below the border-line in intelligence, English and Arithmetic, and her primary teacher perhaps erred on the side of generosity in giving her a forecast of C. In the first two years of the secondary course she clung precariously to her hopeless position of about -1σ, plunged downward in the third, and left school to enter a business college.

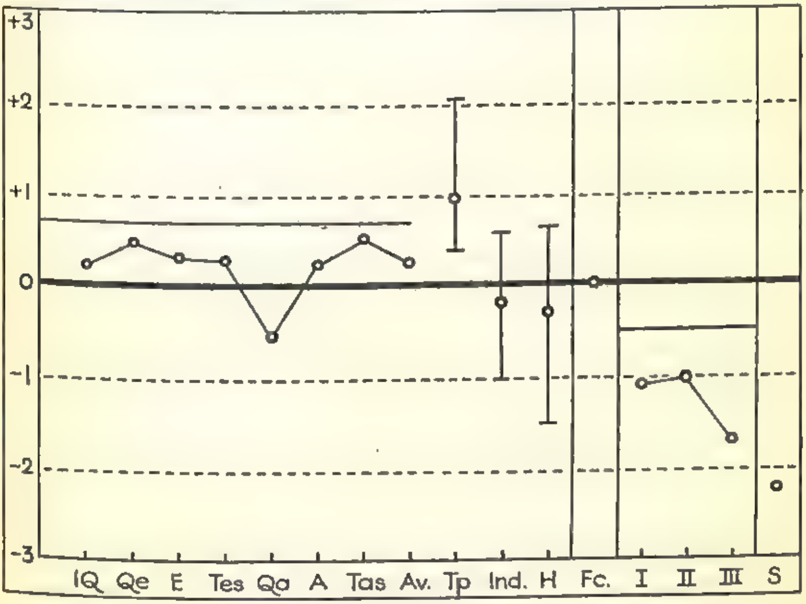

Fig. 12.—A clear reject.

A HOPELESS REJECT

If the parents of this boy had seen his profile (fig. 13) one wonders whether they would have persisted in sending him to a senior secondary school against his own desire. Qualifying at 13 years 9 months, he is weak in everything, has an E grading in industry, and only one marked personal quality, namely, laziness. His later performance amply justifies the primary teacher's pessimistic forecast. The secondary report states that the boy was fit only for a backward class; he had no interest in school-work, and finally started truancy in order to induce his parents to withdraw him.

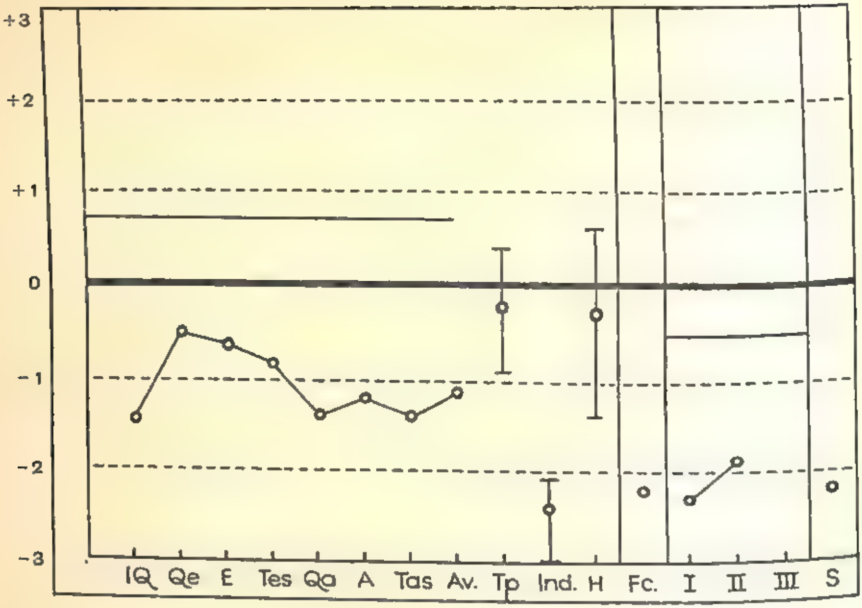

FIG. 13.—A hopeless reject.

These graphs tell six human stories of children very differently endowed. They range from one pupil who, we trust, will be a leader in some department of culture, industry or commerce, to a number who seem destined to be merely hewers of wood and drawers of water. And while they all have their individual educational claims, these were certainly not met by a system under which all but one were allowed to enter a senior secondary school. The admission of pupils with no chance of success is harmful to the pupils themselves, and imposes a heavy burden upon the senior secondary teachers; but there is a third aspect which is sometimes overlooked. In these democratic days we think so much of the claims of the average and dull that we run the risk of forgetting those of the intellectual *élite* who will be the future torchbearers in cultural, social and industrial advance; and it is partly in the interests of this *élite* that improvement in our methods of selection and guidance is necessary.

DISCREPANCIES

It must not be thought that the system is infallible. Later we shall reduce the discrepancies to proper statistical form, but mean-

time we present two cases where the reader, having had his lesson in forecasting, can make his predictions and find himself wrong.

The two pupils in fig. 14 are at very nearly the same standard in the Qualifying data, and both qualified at 12 years of age. If the reader has profited by his lesson, he will have no hesitation in admitting both pupils, confident that they could be clear, but undistinguished, successes in a senior secondary course; if he has

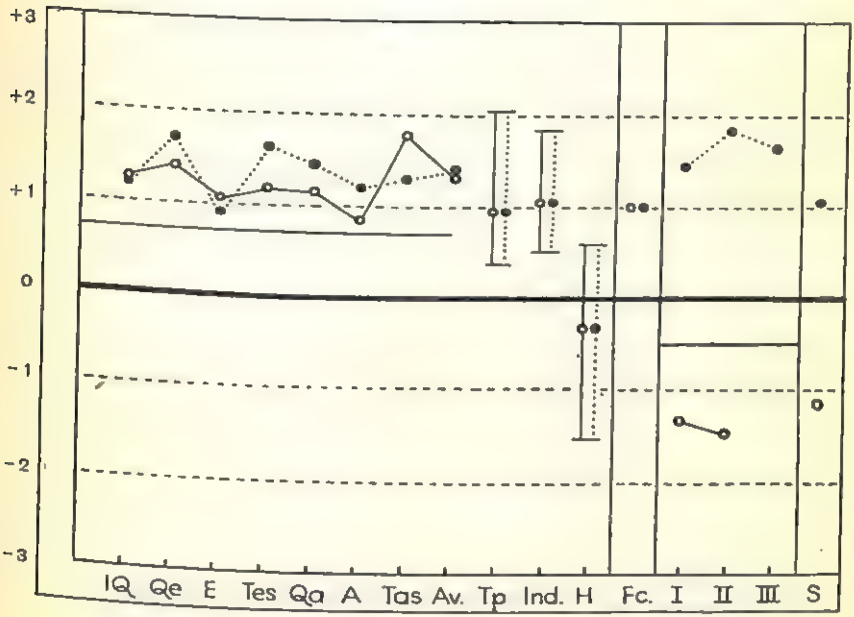

FIG. 14.—Discrepancies between Qualifying score and later success.

given some care to his preparaton he will predict, not only that they would be successes, but that in their success marks they should both have a score of about $+\cdot 3\sigma$.

The right-hand side shows how far he is in error; and his satisfaction that at least one of the pupils is a success is tempered by the fact that this pupil is far more successful than she ought to be. Her success score, instead of being $+\cdot 3\sigma$, is well over $+1\cdot 5\sigma$. In her Qualifying reports she is described as self-reliant and specially talented in Art. In the secondary course she took two foreign languages, French and German, and the head teacher's explanation of her rather remarkable success is that the mother had to bring up the family without help. The pupil was extremely eager to

take full advantage of educational opportunities which, she realised, had been obtained for her with difficulty and sacrifice.

The boy is an entirely different type, yet there was nothing in all our Qualifying records that could have saved us from the mistake of admitting him. He was a clear failure from the start of the secondary course, is reported to have been lazy and inattentive throughout, and left after two years to become a salesman.

TWINS

The two following examples (fig. 15) do more credit to our system. They afford a picture of twin boys who show striking resemblances. Both have exactly the same IQ, the same marked

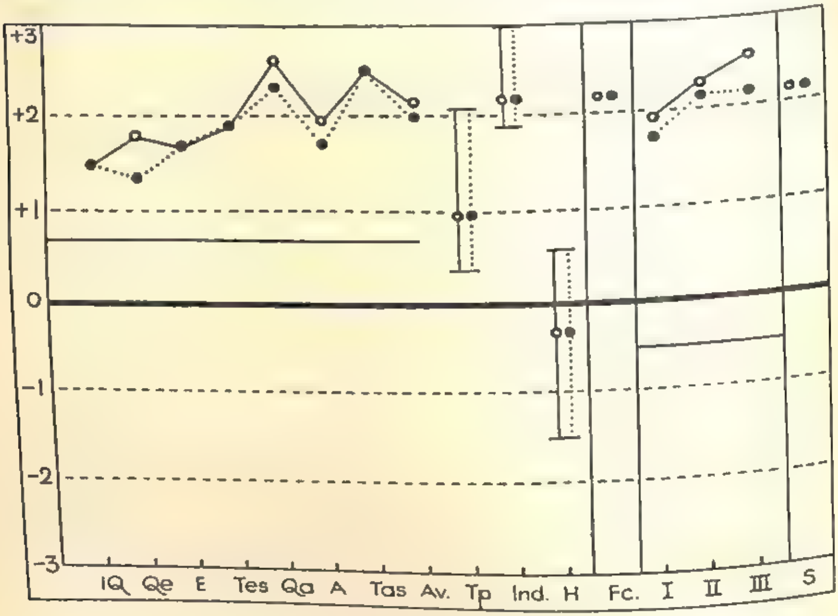

FIG. 15.—Twins.

special talents and the same personal qualities. Prediction was easy here, and proved correct. They are, however, both doing better in the secondary course than would have been expected, even with their high Qualifying average of over +2 sigma. Both obtained their Leaving Certificates with great credit in 1941.

HEALTH

Fig. 16 presents the data for a case of no little difficulty, and is included to emphasise the point that factors other than examination and test results should be considered in selecting and guiding pupils. It is that of a girl of a fine type, with everything in her favour except health. She is a perfectly clear admit, highly intelligent, interested in cultural pursuits, persevering and industrious; a girl of originality and initiative, with special musical

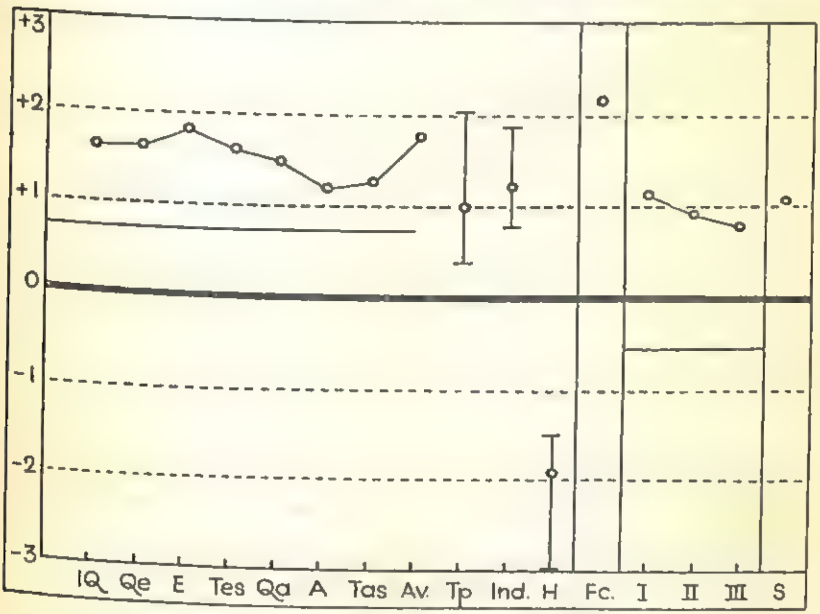

FIG. 16.—An able pupil with a low health grading.

gifts and from a good home. But the Medical Officer's health grading is D, and his confidential report shows that she has heart disease and anæmia. The report from the secondary school mentions that she is 'delicate and often absent.' Despite her physical handicap she has done well in her secondary course, though with distinct signs of falling off.

It is doubtful whether a girl of this type should be encouraged to face the strain of a full Leaving Certificate course, with the temptation to venture on the still more intense strain of an honours degree course at a university.

STEADINESS AND VARIABILITY

Fig. 17 raises a point which we have not yet had time to explore fully. Some pupils are remarkably steady in their performance, others are extremely variable; and it is not always the steady

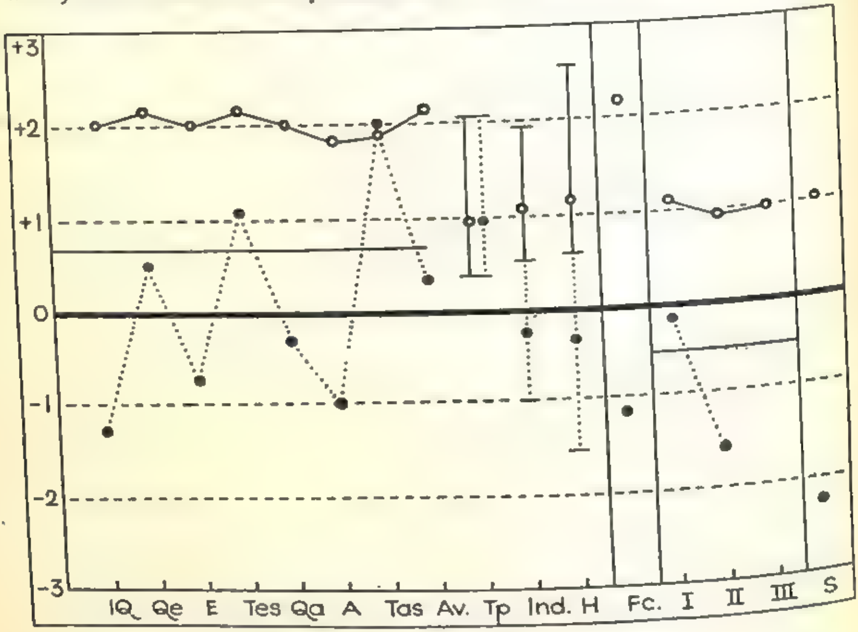

FIG. 17.—Steadiness and variability.

pupil who is the good one and the variable pupil who is the poor one. The variable pupil, however, presents us with the more difficult problem, particularly if he is near the border-line.

SPECIALISED ABILITY

As a rule the pupil who is good in English is good in Arithmetic, but occasionally we come upon instances like the following. The profile (fig. 18) is that of a girl with aspirations to become a music teacher, who has high literary gifts, but is very weak in Arithmetic. This is no doubt partly responsible for her failure in the secondary course, where she found special difficulty with Mathematics. As we shall see later, we should have rejected this pupil for her weak Arithmetic even though she had been considerably above the admission line on the average. This, of course, means merely that in the existing secondary courses she would be unlikely to be

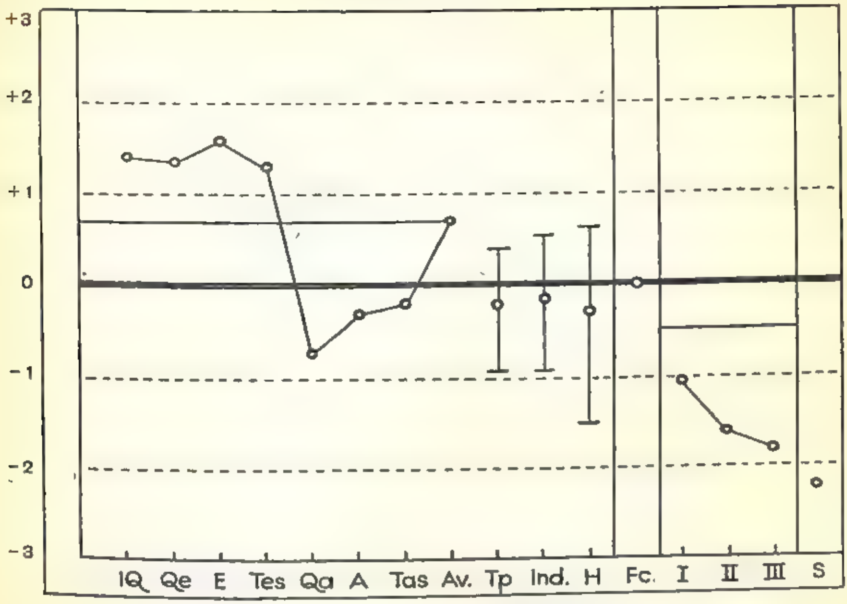

FIG. 18.—A pupil with specialised ability.

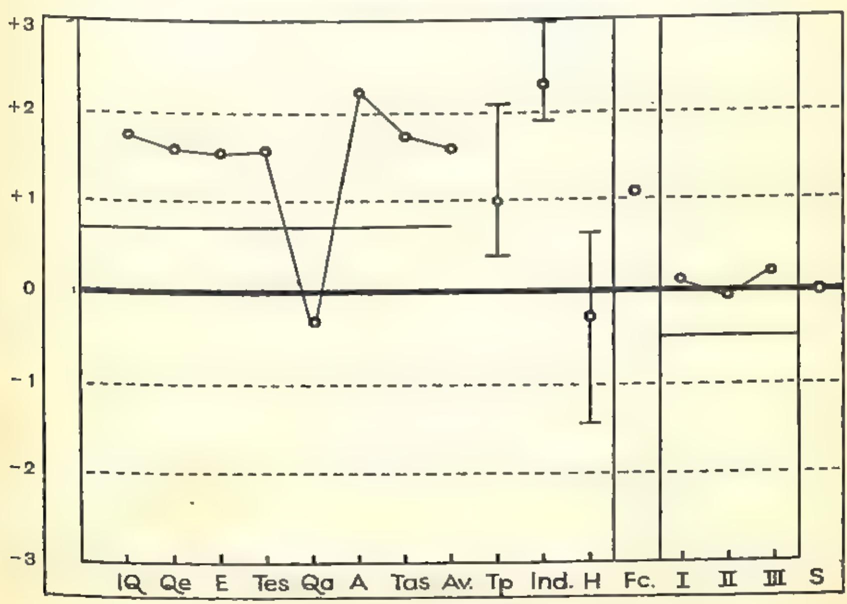

FIG. 19.—A pupil who has made a flop in one subject.

a success. Whether, in a sound educational system, an aspiring musician, with literary gifts of the order of $+1{\cdot}4$ sigma, should be debarred from the chance of attaining her desire because of weakness in Mathematics, is open to question.

A FLOP

Our last illustration (fig. 19) at this point is that of a boy who has made a really thorough flop in his Arithmetic examination. With him no difficulty would arise, for he is a sure admit; but a collapse of this kind in a border-line case might make all the difference between admission and rejection if we had no means of convincing ourselves that the low mark was due to accident. He is not doing quite so well in his secondary course as we should expect, due, according to his head master, to laziness.

CHAPTER XIII

USE OF PROFILES TO DETERMINE THE NUMBER OF MISFITS BY DIFFERENT METHODS OF SELECTION

General Procedure

We know which pupils have turned out to be successful in the secondary course, and which have proved to be failures. If we take the standard average scores on the profiles as the basis of selection, and use a pass-mark of, say, $+1$ sigma, we can now quickly deal the profile cards into two bundles, those whom we should admit and those whom we should reject. These two bundles represent the answers which, on this selection system, we should give to the 452 problems set to us by the applicants for admission to the senior secondary schools. But we now know the correct answers, and we can count our mistakes.

These will fall into two categories: those whom we should have admitted but who have failed (admit-fails), and those whom we should have rejected but who have made good (reject-successes). These two categories together are the misfits.

With this technique the possibilities of rapid experiment are almost limitless. Having ascertained the number of misfits with a pass-mark of $+1$ sigma, we can proceed to make a similar count for the same tests but with a pass-mark of $+\cdot 5$ sigma. We can also deal out the cards on the basis of, say, IQ alone, and find how the number of misfits compares with that for the complete battery.

To the non-mathematical reader this very direct attack on the problem will doubtless carry greater conviction than the correlation results. It will also assist him in his interpretation of correlation coefficients. The statement that $r_{QF} = \cdot 770$ and $r_{IQ\,F} = \cdot 691$ may convey little more than a vague idea that the examination is a better basis of prediction than the intelligence test, leaving him in doubt on two important points:

(1) Whether the difference is an important or a trivial one.
(2) Whether a correlation of ·770 means that the prediction will be virtually perfect, indifferent or good.

He will find it more intelligible if we can tell him the numbers of mistakes which would result from the use of Qualifying examination marks and of intelligence quotient.

SENIOR SECONDARY SCHOOLS

Final Assessment of Success and Failure in the Secondary Course

The use of the above method clearly demands a careful decision on the question whether the pupil has or has not made good in the secondary course; and this was reached in the following way. At the end of the third year of the follow-up each case was scrutinised separately, and we set aside those in which there was no possible doubt. The remainder, including those who had a success grading of C on the follow-up cards, were discussed individually with the head teachers. If, in the opinion of the head teacher, the pupil had more than a 50 per cent. chance of obtaining a minimum Leaving Certificate in a suitable course in six years, he was counted a success; if less than a 50 per cent. chance, he was counted a failure. If a pupil was unlikely to make good through such causes as ill-health or bad home conditions, he was counted a failure even though the head teacher felt that he had the ability to make good under better conditions.

The standard is not high, but we felt that, if a pupil reached it, no Education Committee could honourably exclude him from a senior secondary school. The qualification 'in a suitable course' was intended to meet a not uncommon case where the head master was of the opinion that the pupil was in the wrong course, but that he would have made good if he, or his parents, had made a better choice.

The plan worked extremely well. The cards were arranged in order of merit of the follow-up marks, and at the top and bottom of the list the head teacher usually gave an immediate decision. In between came a doubtful band, much narrower than we had anticipated, where each case was carefully balanced, with considerable reference to mark books and reports. Frequently the head teacher deferred judgment until he had discussed the problem with various members of his staff. For such pupils he was usually

good enough to send a full written report giving detailed reasons for his decision.

One point that emerged clearly from these discussions was that, whatever variation there may be in the standards of marking of the individual schools, there is almost perfect agreement among the senior secondary head teachers as to the border-line which separates those who will gain the Leaving Certificate from those who will not. With the comparable follow-up scores before us we found that, after the interview with the first head teacher, we could tell almost exactly when the others would begin to hesitate, when they would sail into the smooth water of the sure fails, and where they would finally draw the line. We owe a deep debt of gratitude to the secondary head teachers for the courtesy and patient forbearance with which they gave us their help in this difficult task.

Standard Follow-up Score for the Border-line between Success and Failure

When all the decisions had been made we determined, by a method which will be explained later, the point on the scale of standard follow-up scores separating the successes from the failures. It was at $-\cdot 51\sigma$ for the whole group and at almost exactly the same point for each separate school.

Application of the Method of Misfits to test the Value of the Qualifying Examination in Selection for Admission to Senior Secondary Schools

Everything is now in readiness for the application of our post-mortem procedure; and we proceed to show how many mistakes we should have made if we had used the Qualifying examination as our test for admission.

Fig. 20 is a picture whose story should be of considerable interest to all who have the responsibility for selection and guidance at the Qualifying stage, and the non-mathematician will find it more illuminating than the bare statement that $r_{QF} = \cdot 770$.

There is no horizontal scale; so the diagram, while showing whether the pupil was good or bad on the entrance test, does not show the degree of success which he attained in the secondary course—merely whether he was, or was not, a failure. The vertical scale is that of the standard Qualifying examination scores. The pass-mark of the examination is taken as $+\cdot 60$ sigma.

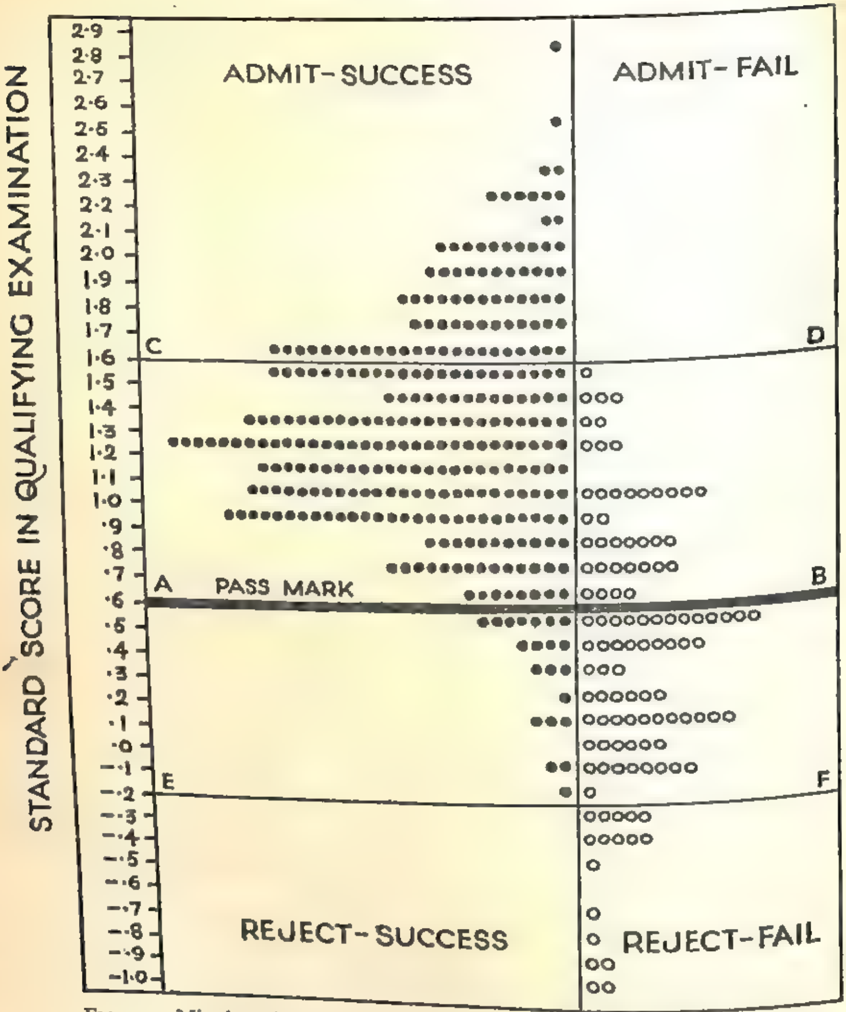

FIG. 20.—Mistakes which would have been made if the Qualifying examination had been used as entrance test for senior secondary courses.

We now count our misfits with the following results:

	Successes	Failures	Total
Admit . . .	283	38	321
Reject . . .	21	74	95
Total .	304	112	416[1]

[1] In order to ensure comparability of the results which follow it was necessary to omit pupils who had absences from any of the seven tests.

Out of the 416 applicants we should have admitted 321, and of these, 38 would not have made good. We should have rejected 95 pupils, 21 of whom would have been wrongly debarred from the opportunity of taking a senior secondary education. The number of mistakes we should have made is 59 out of the total of 416.

Range of Entrance Attainments of the Misfits

One of the most interesting points brought to light by the diagram is the scatter of the misfits on the entrance attainment scale. Pupils with a Qualifying score as high as $+1{\cdot}50$ may be failures; pupils as low as $-{\cdot}10$ may make good. The band of uncertain prediction lies between the lines CD and EF.

Some have aspired to a selection system where no pupils would be admitted to a senior secondary course who would not make good, and are disposed to be critical of the presence of any failures. The diagram shows that it would be possible to satisfy their demands. We should fix the pass-mark at CD; but the purity of the secondary classes would be purchased at the cost of denying a secondary education to no less than 223 pupils fit to profit by it. Others, taking a more humane view, desire to reject no pupils who would ultimately make good. Their demands could be satisfied by lowering the pass-mark to EF, but at the cost of admitting 95 pupils who would clog the educational machine and suffer the depressing effects of early failure.

The Pass-Mark

This leads to the important problem of the correct fixing of the pass-mark, and we use the diagram to introduce the reader to the method by which we studied it. With a pass-mark AB we have 59 misfits. Suppose we raise it to $+{\cdot}70$, we reduce the admit-fails by 4, but we increase the reject-successes by 8, giving a net increase of 4 misfits. On tabulating the results of raising or lowering the pass-mark in this way we come upon an important law.

The law, which we shall find to be general, is that for any entrance test there is a pass-mark at which the number of misfits is a minimum—clearly a law of fundamental importance.

One other notable fact that emerges from a study of Table XXXVII is that, at the point where the misfits are a minimum,

TABLE XXXVII

Effect of changing the pass-mark on the actual number of misfits (Entrance test, Q)

Pass-mark	Admit-fail	Reject-success	Misfits
+1·60 sigma	0	223	223
+1·40	4	186	190
+1·20	9	130	139
+1·00	18	81	99
+ ·80	27	43	70
+ ·70	34	29	63
+ ·60	38	21	59
+ ·50	51	14	65
+ ·40	60	10	70
+ ·20	69	6	75
·00	86	3	89
− ·20	95	0	95

the number of admit-fails is considerably greater than the number of reject-successes.

The Correct Pass-Mark

It will now be convenient to consider the general question of the position of the pass-line. This is a problem on which science can throw some light, but the complete solution transcends scientific facts. Our finding depends to some extent on whether we think it does more harm to reject a success than to admit a failure. If the philosopher will solve that conundrum, and put his answer in mathematical terms (for example, that the rejection of a success is twice as harmful as the admission of a failure), the scientist will do the rest.

As we know of no authoritative pronouncement on the point, we shall assume that the two types of mistake are equally harmful. That being so, we suggest that the pass-mark is drawn correctly when the number of misfits is at the minimum.

We shall find later that the minimum always comes at a definite point, at which:

The pupil who is admitted has more than a 50 per cent. chance of succeeding; the pupil who is rejected has less than a 50 per cent. chance of succeeding.

In all cases likely to arise in Scotland, the number of admit-fails will be greater than the number of reject-successes.

If we err, therefore, it is on the side of giving the child his educational chance.

Another possible pass-mark is the point at which the number of admit-fails is equal to the number of reject-successes. This has a specious air of fairness, but a glance at Table XXXVII will show that it would give a larger number of misfits, and its unsoundness will be more clearly disclosed later.

We shall in future always refer to the point at which the number of misfits is a minimum and the child has just an even chance of success as the *correct pass-mark.*

We now proceed to compare the predictive value of the various tests, examinations and batteries by finding the number of misfits each would give if it were used with its own correct pass-mark.

Method of Tabulation

Tables of the following form were prepared for all groups, senior and junior secondary, backward and retained.

Class interval	Senior Leaving Certificate		Junior Leaving Certificate		Day School Certificate (Lower)		Shorter post-primary course		Backward class	
	S*	F†	S	F	S	F	S	F	S	F
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
+·40–·49 (sigma)	4	17	7	14	21	0	21	0	0	0
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

* S = Success † F = Failure

In the case of the Junior Leaving Certificate a pupil was graded a success if, in the opinion of the head teacher, he had more than a 50 per cent. chance of obtaining the Certificate in three years. A similar test was applied for the Day School Certificate (Lower), the time allowed being two years.

Determination of the Minimum Number of Misfits for a Test or Battery

There are two difficulties in determining the minimum number of misfits. In the first place, the minimum may lie within a class interval; in the second place, the numbers within class intervals of ·1 sigma are small, and there is consequently a degree of unsteadiness due to the chance fluctuations of small numbers. We must therefore try to introduce some of the steadying effect of the numbers above and below the pass-line.

If we take Table XXXVII as our illustration, the actual number of misfits is 59 with a pass-mark of +·60 sigma. To get a fairer estimate of the general tendency of the Table we could plot the number of misfits against the pass-marks, smooth the curve, and find where the smooth curve marks the minimum.

Generally speaking, there are two ways of smoothing curves: mathematical methods, like the method of least squares, and graphical methods. The first method is the soundest but involves very heavy calculations. The second method can be carried out rapidly without calculation, and works very satisfactorily with certain types of curve. The essence of such methods is that we replace each point on the curve by the weighted average of it and of the points above and below.[1]

The disadvantage of the graphical method is that it tends to make a concave curve a little less concave, and all the curves of misfits are of this type. To avoid this we tried various expedients, but finally decided to plot separately the curves for admit-fails and reject-successes. These, as will be seen from fig. 21, were much more suitable for graphical smoothing.

The figure shows how the smoothing removes the unsteadiness which comes from small numbers. The number of misfits for any pass-mark is found by reading from the graphs the numbers of admit-fails and reject-successes, and adding.

Determination of the Correct Pass-Mark

The correct pass-mark was determined in two ways. At the correct pass-mark the number of misfits is a minimum; therefore

[1] If we have a series of points $y_1, y_2, y_3, \ldots y_n$, to find the smoothed point corresponding to y_2, a ruler is placed through y_1 and y_3. The mid-point of the vertical distance between y_2 and the ruler gives the smoothed point. In this way we replace y_2 by $\frac{y_1+2y_2+y_3}{4}$. In refractory cases the process can be repeated with the smoothed points.

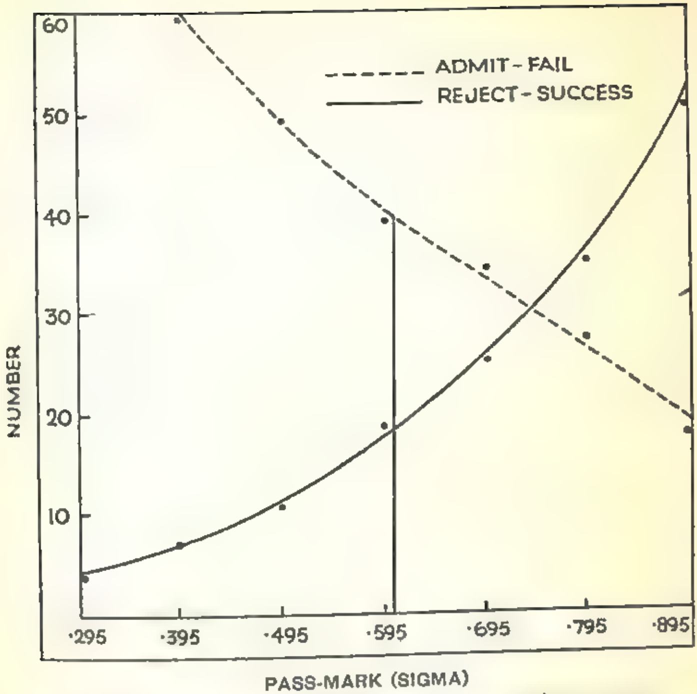

Fig. 21.—Smoothing of curves of admit-fails and reject-successes (Battery IQ + S + Ts).

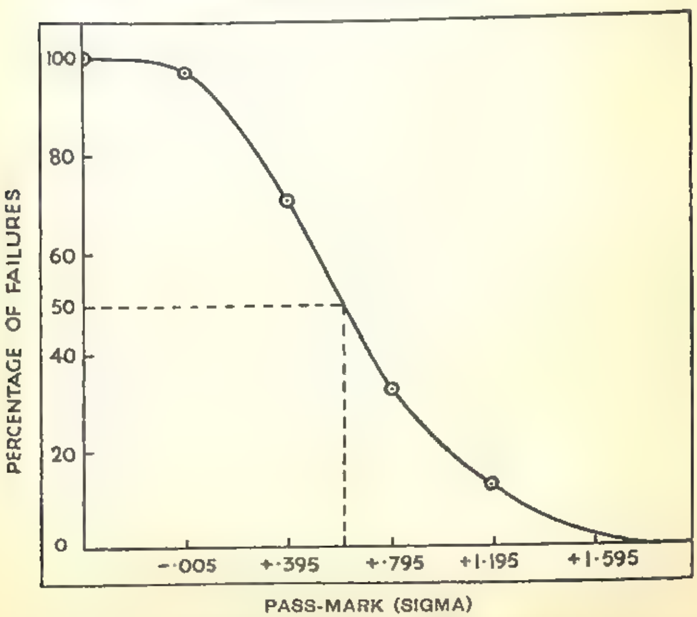

Fig. 22.—Determination of correct pass-mark from percentage of failures (Battery IQ + S + Ts).

we can find it from the above-mentioned graphs. The pupil has an even chance of success; and the correct pass-mark will be the point at which the 50 per cent. line cuts the curve formed by plotting the percentage of failures for each class interval. This is shown in fig. 22.

Figs. 21 and 22 are both in respect of the battery IQ + S + Ts, and by both methods the correct pass-mark comes out to be +·673 sigma.[1]

Correct Pass-Marks for Various Tests and Batteries

The pass-marks obtained in these two ways for all tests and batteries are given in Table XXXVIII. The batteries are arranged in order of the correlation coefficients, which are given in the first column.

TABLE XXXVIII

Correct pass-marks in terms of sigma for all tests and batteries

Battery	Correlation with F (1)	Pass-mark (50 per cent. chance) (2)	Pass-mark (minimum misfits) (3)	*d* (4)	Average pass-mark (5)	Average pass-mark smoothed (6)	Theoretical pass-mark (7)
IQ+Q+Ts	·804	·696	·707	−·011	·701	·70	·68
IQ+Q+S+Ts	·800	·700	..	..	·700	·70	·67
Q +S +Ts	·790	·701	·701	·000	·701	·69	·66
IQ+Q	·786	·676	·644	+·032	·660	·69	·66
Q +Ts	·783	·680	·713	−·033	·696	·68	·66
IQ+Q+S	·783	·654	·676	−·022	·665	·68	·66
IQ+S +Ts	·782	·673	·673	·000	·673	·68	·66
IQ+Ts	·779	·680	·646	+·034	·663	·67	·65
Q +S	·774	·660	·649	+·011	·654	·66	·65
Q	·770	·631	·631	·000	·631	·66	·65
S +Ts	·764	·666	·666	·000	·666	·65	·64
IQ+ S	·736	·572	·647	−·075	·609	·62	·61
Ts	·720	·642	·566	+·076	·604	·61	·59
S	·698	·633	·666	−·033	·649	·58	·56
IQ	·691	·475	·475	·000	·475	·57	·55

The discrepancies between the two methods of finding the pass-mark, shown in column (4), are not great, and exceed ·034 sigma only for IQ + S and for Ts. To get a more reliable estimate

[1] In determining the correct pass-mark it has to be kept in mind that the standard deviation of IQ + Q + Ts is less than unity.

of the pass-mark the two measures were averaged, giving the results in column (5).

THE HIGHER THE FOLLOW-UP CORRELATION FOR THE BATTERY THE HIGHER THE CORRECT PASS-MARK

Column (5) reveals the important law that the higher the predictive value of the battery the higher should be its pass-mark. In the general discussion, to which this chapter is an empirical

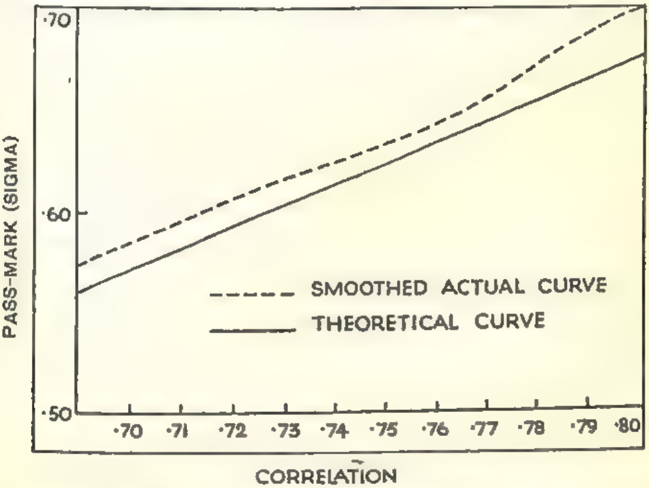

FIG. 23.—Fall of pass-mark with fall of correlation: theoretical curve and smoothed actual curve.

preamble, the reason for this will be given, and we shall find that the theoretical curve of fall of pass-mark with fall of correlation is smooth. We may therefore assume that we should get more reliable estimates of the pass-marks by smoothing the actual curve. The result of the process is shown in fig. 23.

No special significance attaches to the fact that the theoretical curve is a little lower than the actual one: this is due to the skewness of the distribution of the senior secondary group. It is required at the moment merely to act as a check on the general sweep of the actual curve; and when we consider the size of the vertical scale it is clear that the agreement is satisfactory. We therefore take the points obtained from this smoothed curve as the best estimates we can get of the pass-marks for the batteries; and these are given in column (6) of Table XXXVIII. The pass-marks obtained from the theoretical curve are given in column (7).

For assessing the significance of the small possible errors in the determination of the pass-marks we give the following guide:

An error of ·03 sigma in fixing the pass-mark would increase the misfits by about 1 per cent.

An error of ·02 sigma would increase the misfits by about 1 in 200.

An error of ·01 sigma would increase the misfits by about 1 in 500.

Number of Misfits for the Various Tests and Batteries

We now come to the misfits for the various batteries, full details of which are given in Table XXXIX. For the sake of those who regard all statistical manipulation with suspicion we give, in column (2), the actual minimum misfits obtained from the distributions with a class interval of ·1 sigma. These, however, are not the exact minima, which would usually fall *within* a class interval. In columns (3), (4) and (5) we give the minima obtained as explained above from the smoothed curves.

TABLE XXXIX

Number of misfits for the various tests and batteries

Battery	Correlation with F	Actual minimum misfits	Actual minimum misfits (smoothed)			Minimum misfits for normal correlation			Correlation corresponding to actual misfits (col. (5))
			Admit-fail	Reject-success	Misfits	Admit-fail	Reject-success	Misfits	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IQ+Q+Ts	·804	56	38·1	18·4	56·5	44·0	24·4	68·4	·802
IQ+Q+S+Ts	·800	55	37·1	19·4	56·5	44·5	24·7	69·2	·802
Q+S+Ts	·790	56	37·4	20·0	57·4	46·0	24·8	70·8	·793
IQ+Q	·786	55	38·5	19·0	57·5	46·7	24·9	71·6	·792
Q+Ts	·783	59	38·9	20·1	59·0	46·9	24·9	71·8	·780
IQ+Q+S	·783	58	37·2	21·9	59·1	46·9	24·9	71·8	·779
IQ+S+Ts	·782	58	39·5	19·2	58·7	47·0	24·9	71·9	·782
IQ+Ts	·779	55	41·1	17·0	58·1	47·5	25·0	72·5	·787
Q+S	·774	57	40·2	19·1	59·3	48·1	25·0	73·1	·777
Q	·770	59	41·1	20·4	61·5	48·9	25·1	74·0	·760
S+Ts	·764	56	40·2	20·1	60·3	49·6	25·2	74·8	·770
IQ+S	·736	62	39·0	25·1	64·1	53·7	25·5	79·2	·736
Ts	·720	59	48·0	13·0	61·1	55·8	25·4	81·2	·763
S	·698	64	40·6	27·4	68·0	58·6	25·4	84·0	·701
IQ	·691	69	47·1	22·2	69·3	59·6	25·3	84·9	·690

There is no serious discrepancy between the actual minima and those obtained from the smoothed curves except perhaps in the case of S + Ts, where a study of the curves leaves no doubt as to the greater soundness of the smoothed result. The way in which these deviations arise may be explained by reference to fig. 21. There the minimum is indicated by a vertical line near which the actual point on the admit-fail curve lies below the smooth curve. For the reject-success curve the opposite holds. Here, therefore, the actual minimum would be close to the smoothed minimum; but if the actual points had fallen below the smoothed curves in both cases the discrepancies would have been additive, and the difference considerable.

The numbers of misfits range from 56·5 for the best battery to 69·3 for the poorest. These should be considered in relation to 112, which is the number of misfits we should have had if we had dispensed with the entrance examination and admitted all the candidates. Our poorest battery gives a correlation with F of ·691. By methods described in Chapter XIV we can estimate the numbers of misfits for batteries giving lower correlations. These are as follows:

Correlation with F	Minimum number of misfits
·6	82
·5	94
·4	105
·3	111
·2	112
·1	112
0·0	112

From these it will be noted that entrance examinations giving a correlation with F of less than ·4 are of very little value.

Orders of Predictive Value based on Correlations and on Misfits

The relationship between the two orders will be shown later in graphical form. Meantime, an inspection of columns (1) and (5) of Table XXXIX will show that the conformity between the two is fairly close. The first six batteries are in the same order on both bases; and the differences which cause the later discrepancies in order are small.

It is, of course, not necessary that the orders should be the

same. The correlations are based on the whole of the data, and give the order of discriminative capacity of the batteries throughout the scale: the misfits give the order of discrimination at the border-line. As far as practical application is concerned we might put the position in this way. The order of misfits is the more significant for the Education Committee which is

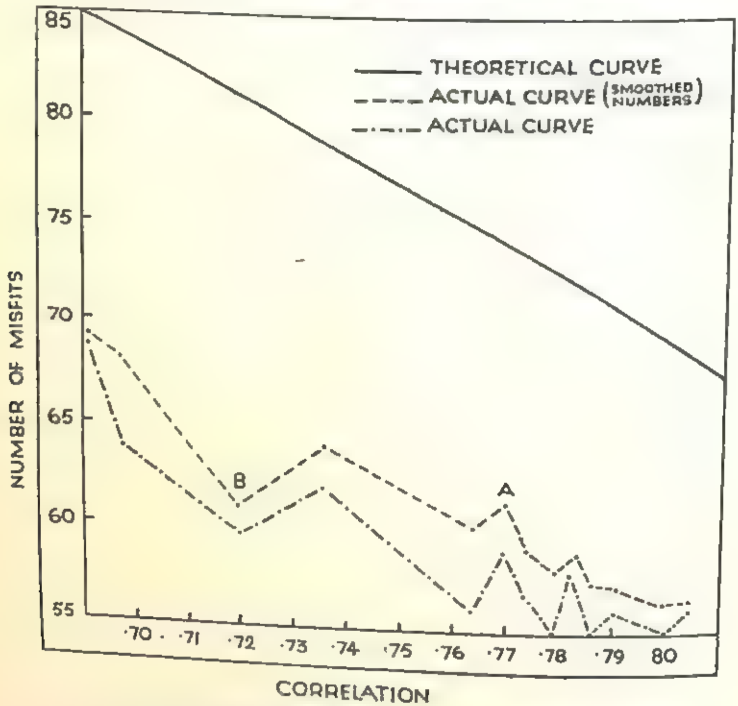

FIG. 24.—Comparison of actual and theoretical number of misfits for different sizes of correlation.

concerned merely with the selection of pupils for admission to senior secondary schools: the order of correlations is the more significant for the secondary head master who will use the results of the entrance tests for arranging his entrants into homogeneous groups.

In Chapter XIV we shall see that it is possible to calculate theoretically the number of misfits for any size of normal correlation. For comparison, these theoretical values are given in columns (6), (7) and (8) of Table XXXIX, but the relationship is best shown by the graphs in fig. 24.

The fluctuations are somewhat less for the smoothed actual curve than for the actual curve; but they are far from completely eliminated, and give the impression that the deviations are real and not chance variations. The point A corresponds to Q, and B to Ts; and while Q has the higher correlation, Ts gives the fewer misfits. It may well be that this is no mere accident. Q may get its high correlation by properly placing the better pupils, while the teachers' estimates may be particularly discriminating at the border-line.

Comparison between Actual and Theoretical Numbers of Misfits

The theoretical and the smoothed actual curves differ in two ways. In the first place, the actual numbers are all much less than the theoretical. This is no doubt due, in small part, to the under-estimation of the correlations through a cumulation of small losses, but mainly to the fact that the distribution of the senior secondary group is tailed. In the second place, the actual curve has a slightly smaller slope than the theoretical. This is related to the first difference, as can be seen if we imagine the curves completed to the ends, $r = 0$ and $r = 1$.

Correlations which would give the Actual Numbers of Misfits

Most of our data are in the form of correlations, and ultimately we shall gather together all our evidence in order to get a final order of predictive value for the various tests and batteries. It would therefore be convenient if we could put the order based on misfits in this comparable form. Since the theoretical and the actual numbers of misfits are not the same, we cannot do so by finding the normal correlations that would give the number of misfits for each battery. The only method is to draw a smooth curve through the points for the actual smoothed curve and read the correlations from it. In so doing we are not assuming, contrary to what has been said above, that the actual smoothed curve should have the same smooth sweep as the theoretical one. We retain the order of the misfits and simply put the data into another form comparable with correlations. These fictitious correlations are given in column (9) of Table XXXIX.

Composition of Misfits

From Table XXXIX it will be seen that there are considerable fluctuations in the proportions of admit-fails and reject-successes. This is due to the fact that while the total number of misfits is not very sensitive to small changes in the placing of the border-line, the proportions are. The numbers in the Table should therefore be taken as giving merely a rough indication of the proportions.

JUNIOR SECONDARY SCHOOLS

Distribution of the Junior Secondary Group

The fact that the distribution of the senior secondary pupils is tailed caused certain differences between the theoretical and the actual results. As will be seen from fig. 57,[1] the distribution of the junior secondary group is more nearly normal and we should anticipate that the correspondence would be closer.

Qualifying Pass-Mark for Success in the Junior Leaving Certificate

As for the senior secondary courses, we determined the point on the standard scores for IQ + Q + S + Ts at which the pupil has a 50 per cent. chance of success in the Junior Leaving Certificate. For this battery the correct Qualifying pass-mark is as follows:

Pass-mark (50 per cent. chance)	Pass-mark (minimum misfits)	Theoretical pass-mark
+ ·595 sigma	+ ·595 sigma	+ ·597 sigma

The border-line for success in the Junior Certificate is only about ·1 sigma lower than that for the Senior Certificate. The standard for the Certificate is therefore fairly high.

Should anyone wish to find the correct pass-mark for one of the poorer batteries he will get a very close approximation by subtracting ·1 sigma from the appropriate number in column (6) of Table XXXVIII.

[1] P. 203.

Number and Composition of Misfits in Selection of Pupils for Junior Leaving Certificate

Using the methods already described, we can find the number of mistakes we should make if we used the battery IQ + Q + S + Ts, with its correct pass-mark, to select pupils who should take a Junior Leaving Certificate course in junior secondary schools. The total number in our junior secondary group was 1,975.

	Admit-fails	Reject-successes	Misfits
Actual . .	164	258	422
Theoretical . .	143	276	419

CHAPTER XIV

GENERAL PRINCIPLES OF SELECTION TECHNIQUE

We have, by empirical procedure, come upon several laws governing the operation of the selection process. Since these might be specific to our data, the time is ripe to gather the various points together in a theoretical treatment which will show how far they are general.

The present chapter is an attempt to systematise a collection of principles which we encountered from time to time in working up the Inquiry data. Starting from peculiarities of our results we made various explorations into the field of theory, and found that a surprisingly large number of laws could be deduced mathematically without experiment. We felt that our own picture of the selection situation was clarified in this way, and we hope that this chapter may do the same for our readers. Certainly it brings out principles which should be in the minds of all responsible for selecting pupils for secondary education or awarding bursaries or free places.

Our honesty in recommending the non-mathematical reader to omit Chapter VI may give him greater confidence in our advice that he should read this one. He should not be deterred by the fact that, at places, it has a mathematical appearance; this is deceptive. A certain amount of difficult mathematics undoubtedly is involved in some of the operations; but we believe that, by taking the results of these operations for granted, he will be able to follow the drift of the reasoning and understand how the various principles are reached.

The Scatter Diagram

Fig. 20[1] shows whether the pupil is a high or a low admit, but gives no indication whether he was a distinguished or merely a border-line success in the secondary school. We can display the results in a way that will give us both pieces of information by a figure of the following type:

[1] P. 106.

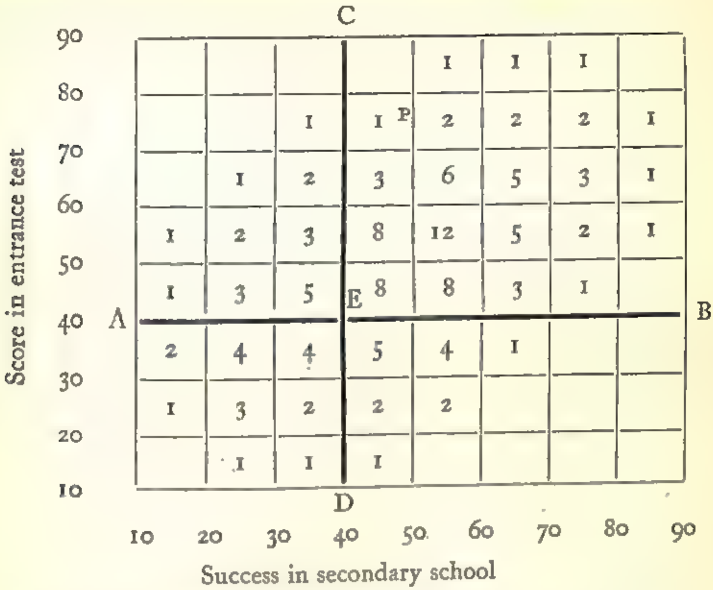

FIG. 25.—Scatter diagram showing position of pupil both in entrance test and in success.

The pass-mark on the entrance test will be represented by a horizontal line like AB, and the line separating successes from failures in the secondary course by a vertical line like CD. Pupils in the quadrant CEB will be admit-successes; those in the quadrant BED reject-successes, and so on. If a pupil's position is in cell P we know that his entrance test score was between 70 and 80 per cent., and that his success score was only between 40 and 50 per cent.

It is from scatter diagrams of this kind that correlations are calculated, and the entries usually fall within an ellipse. A long narrow ellipse means that a pupil with a high score on the entrance test generally has a high score in success, and *vice versa*. The correlation is therefore high. A broad ellipse, approaching circular shape, indicates a low correlation; a long narrow ellipse lying along the 90–90 diagonal would signify a high negative correlation.

In any row or in any column the numbers are high in the middle and decrease towards the ends.

The Normal Bivariate Surface

The title is forbidding, but the notion is simple. If we were to erect a square tower on each cell, the height of the tower being proportionate to the number of the pupils in the cell, we should get a solid which is high towards the centre and lower towards the sides and ends; and if we imagine the numbers to be increased and the class intervals diminished, we should in the

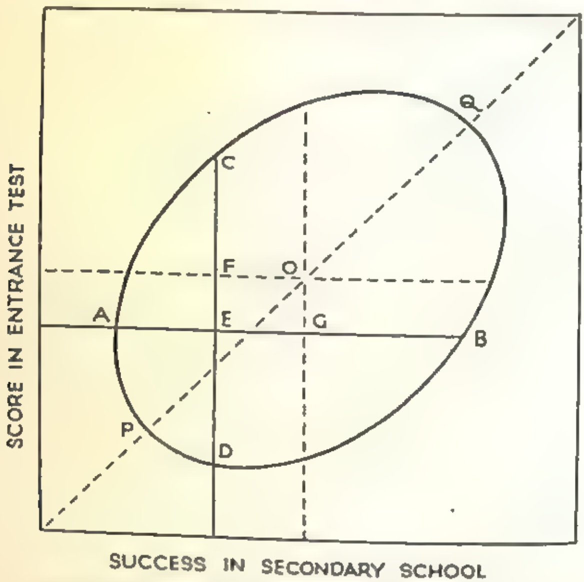

Fig. 26.—Schematic representation of the bivariate surface corresponding to the correlation between entrance test and success.

end get a smooth surface resembling that of a policeman's helmet. If the distributions in the rows and columns are normal, the result is called a normal bivariate surface, and it has the following properties. Each vertical section, whether parallel to the x- or the y-axis, is a normal distribution curve, and each horizontal section is an ellipse.

In view of the latter property we can use an ellipse as a shorthand representation of the bivariate surface for any particular correlation. If the ellipse were at a distance of about 3σ from the centre, it would include practically all the cases, and might be regarded as the base of the solid. Our standard figure for further discussion is thus of the type shown in fig. 26.

When the correlation has a bivariate surface with the above-mentioned properties we speak of it as a *normal correlation.*

AB is the pass-mark on the entrance test; CD the line dividing successes from failures in the secondary course; O is the centre of the ellipse and PQ the major axis.

Basis of the Mathematical Treatment

Our problems now become amenable to mathematical treatment, for the equation of the normal bivariate surface is known.

If we refer to the scatter diagram in fig. 25 we see that the number of pupils in the quadrant CEB is the volume of the solid cut off by vertical planes through CE and EB. Similarly, the number of misfits is the sum of the volumes CEA and DEB. Since the equation of the surface is known these volumes can be determined mathematically.[1]

Problems of Selection and Guidance

There are three main problems at the threshold of secondary education:

- Prediction of the success of an individual pupil in the secondary course.
- Selection of pupils for a definitely limited number of scholarships or free places.
- Selection of pupils who are fit for secondary education, the number not being definitely limited.

PREDICTING THE SUCCESS OF AN INDIVIDUAL PUPIL

To make the problem more definite we take two sample figures. The average correlation between an ordinary entrance examination and success in the secondary school in four recent careful investigations (three in England and one in Australia) is $+\cdot 34$.

[1] If OF expressed in terms of the standard deviation of $x=h$, and OG expressed in terms of the standard deviation of $y=k$, then the proportion of misfits is given by

$$\frac{1}{2\pi\sqrt{1-r^2}}\left\{\int_h^{-\infty}\int_k^{\infty} e^{\frac{-(x^2-2rxy+y^2)}{2(1-r^2)}}\,dxdy+\int_h^{\infty}\int_k^{-\infty} e^{\frac{-(x^2-2rxy+y^2)}{2(1-r^2)}}\,dxdy\right\}$$

References: G. Udny Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, Chapter xii.

K. Pearson, *Tables for Statisticians and Biometricians*, Part II. First edition. London: Cambridge University Press. Mathematical Introduction, p. lii.

Our correlation for the best battery, one of the highest ever published, is +·80. With what assurance could we, in these two cases, predict a pupil's success if we were given his score at entrance?

The methods of dealing with this problem are well known and are dealt with here mainly for the sake of comparison with what follows in regard to group selection.[1] They are based upon the fact that if the correlation between two variates is known we can

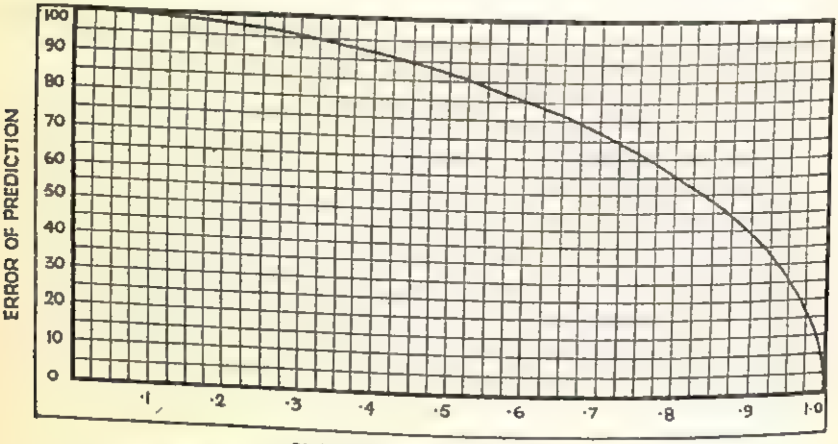

FIG. 27.—Significance of different sizes of correlation in predicting the success of an individual pupil.

find the standard error of an estimate of the second when given the first.[2]

Clearly, if the correlation were +1, the error of our estimate of a pupil's success, given his entrance test score, would be zero. If the error of estimate by pure guesswork[3] were 100, we can, by this method, give the error of estimate for any size of correlation in the form of a percentage. This is done graphically in fig. 27.

We can now find the answers to our two questions. If the correlation between entrance examination and success is +·34,

[1] T. L. Kelley, *Statistical Method*, Chapter viii; *Interpretation of Educational Measurements*. New York: World Book Company, 1927. Chapter vii. A. S. Otis, *Statistical Method in Educational Measurement*. New York: World Book Company, 1925. Chapter xvii.

[2] $\sigma_{1\cdot2} = \sigma_1\sqrt{1 - r^2_{12}}$.

[3] By 'pure guesswork' we mean that we predict with no knowledge of the pupil's entrance test score, or on the basis of an entrance test whose correlation with success is zero. In both cases we would predict the average of the group.

the error of prediction of the success of an individual pupil would be about 94 per cent. of the error by pure guesswork. Even with our own highest correlation of ·8 it would be 60 per cent.

These results show the uncertainty of any guidance given to a pupil solely on the basis of examinations and tests, no matter how complete or carefully conducted.

SELECTION FOR SCHOLARSHIPS AND FREE PLACES

Here we are interested, not in the individual pupil, but in the number of wrong awards which we should make with different systems of selection.

A Sample Case: Award of 100 Scholarships with 365 Candidates

We use this as an illustration of the mathematical method; and we shall show how, if we know the correlation,[1] we can tell how many wrong awards we should make.

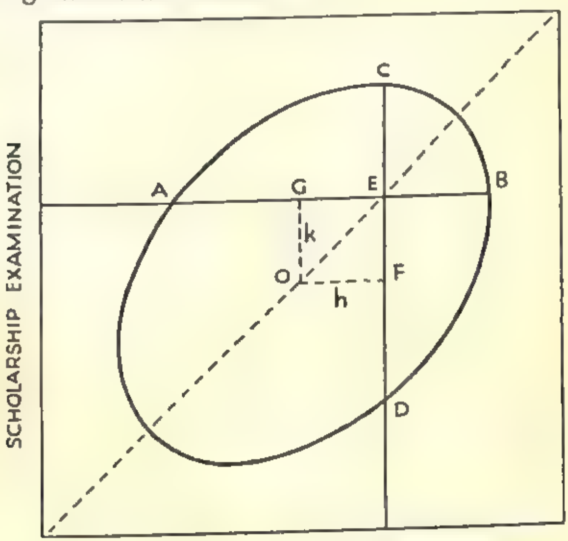

FIG. 28.—Award of 100 scholarships with 365 candidates.

The pass-mark AB is fixed for us because it must cut off exactly 100 applicants. Since we are assuming normal correlation, we can easily find how far AB must be above the mean of the examination

[1] At this and various other points in the Report we have to bear in mind the distinction between a correlation based upon a truncated distribution and the correlation based upon the complete distribution.

distribution if it is to cut off this number.[1] This distance, OG, in terms of the standard deviation of the distribution, is denoted by k. In this case $OG = k = +\cdot 6\sigma$. The volume ACB then indicates the pupils who actually get scholarships.

Since these scholarships are awarded for secondary education, we may assume that the intention is to award them to the 100 pupils who would be most successful in the secondary course. The line dividing those who should have got scholarships from those who should not is therefore a vertical line CD which cuts off the top 100 as far as success is concerned. This must be $\cdot 6\sigma$ above the mean, so that $OF = h = +\cdot 6\sigma$. AB and CD must meet on the major axis of the ellipse, and the volumes AEC and BED must be equal.

We can now interpret our picture:

The pupils in volume CEB got scholarships, and they come into the top 100 in success. They therefore merited their awards.

The pupils in volume AED were correctly refused scholarships.

The number of pupils who got scholarships in error is the volume AEC.

The number of pupils who were unjustly refused scholarships is the volume BED.

The number of mistakes which we should have made is the sum of the volumes AEC and BED. Given the size of the correlation it is possible to calculate all the volumes, as already explained, with the following results:

TABLE XL

Award of 100 scholarships with 365 candidates

Method of award	Correct awards	Wrong awards	Wrong refusals	Total mistakes
Pulling names out of a hat *	27	73	73	145
Examination :				
$r = +\cdot 34$ . . .	42	58	58	116
$r = +\cdot 50$ . . .	50	50	50	100
$r = +\cdot 80$ . . .	69	31	31	62

* The numbers in this row are the *most probable* numbers.

[1] This is done by means of Tables of the Probability Integral. *Cf.* K. Pearson, *Tables for Statisticians and Biometricians*, Part I. Table II.

These are again disappointing results, particularly when one considers that few scholarship examinations, judging from the published results, are likely to have a correlation with success as high as ·5. They will bring comfort only to those who at some time have been unsuccessful in such competitions.

Curve showing the Significance of Correlations of Different Sizes in the award of Scholarships

In dealing with the general problem it would be interesting, and sometimes helpful in practice, to have a curve similar to that in fig. 27 which would show the fall in the number of mistakes with increase of correlation between the scholarship examination and success.

In making such a curve there is this difficulty, that the number of mistakes depends not merely on the size of the correlation but also on the ratio of the number of scholarships to the number of candidates. We require therefore not one curve but many—one, indeed, for each ratio.

Given the ratio it is easy to make the curve. The ends are fixed. If $r = +1$, the number of mistakes is zero; if $r = 0$, the number is the same as we should get by chance selection. The intermediate values can be calculated by the method described above.[1]

The results for different ratios of scholarships to candidates are shown in fig. 29.

The curves show that the predictive value of a correlation coefficient rises most rapidly when the number of scholarships is half the number of candidates. In this case the number of mistakes is reduced by one-half when the correlation is about ·7. If the number of scholarships is very small, say ·5 per cent. of the number of candidates, it would take a correlation of over ·9 to effect the same reduction.

For practical purposes a nest of different curves is not very convenient, but we can make a simplification in the following way. By reference to the symmetry of fig. 28 we see that the number of mistakes would be the same when $h = k = -{\cdot}6\sigma$ as it is when $h = k = +{\cdot}6\sigma$. Further, the number of scholarships or bursaries awarded is not likely, at any rate at the Qualifying stage in Scotland, to be less than 16 per cent., or more than 84 per cent., of the total number of candidates. All cases between these limits fall

[1] When the number of scholarships is half the total number of candidates there is a simpler method. In this case proportion of mistakes $= \frac{1}{\pi} \cos^{-1} r$.

within the narrow band enclosed by the two lower curves. A curve drawn midway between them will therefore give a fairly

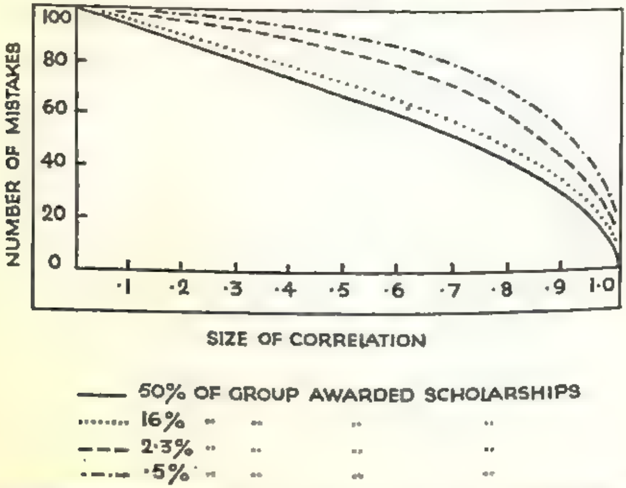

FIG. 29.—Significance of different sizes of correlation n award of scholarships: Exact curves for different percentages.

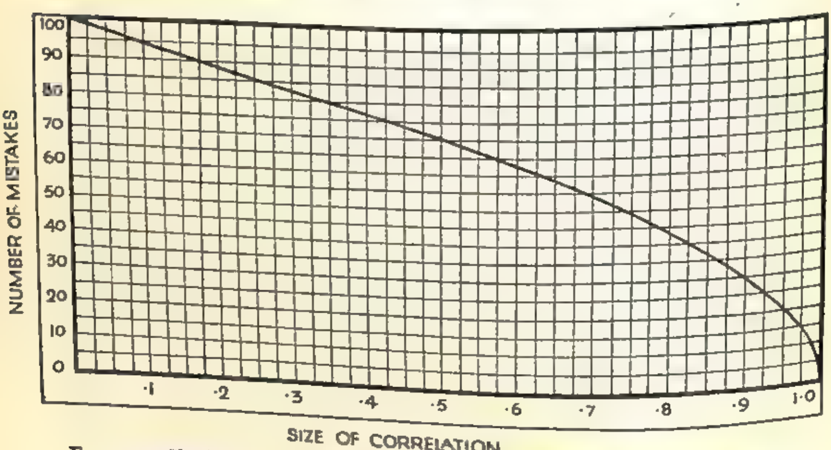

FIG. 30.—Significance of different sizes of correlation in award of scholarships: Approximate single curve.

close approximation to the number of mistakes for all the cases likely to arise in practice.

The curve determined in this way is shown in fig. 30. It is carefully drawn, and can be used in practice to find the number of mistakes for any given size of correlation.

Example of the Use of the Curve in fig. 30

To illustrate the way in which the curve could be used we take the sample case given above, where we know the correct answer. Here we have to award 100 scholarships with 365 candidates.

The first step is to find the number of mistakes by pure chance selection. This is easily done since we know that, among the first 100 drawn from the hat, the number of incorrect awards will be $\frac{265}{365} \times 100 = 72{\cdot}6$. The number of mistakes will therefore be 145.

If, now, the scholarship examination gives a correlation of ·5 with success, we find from the graph that the number of mistakes will be 70 per cent. of the number by pure chance, that is, 70 per cent. of 145. This is 101. We shall therefore give scholarships to 50 pupils who should not receive them, and refuse scholarships to 50 pupils who deserve them. Also we know that out of the 100, only 50 will be correct awards.

These figures correspond very closely with the exact results given in Table XL. If, therefore, a Director of Education knows approximately the correlation between a scholarship examination and success, he can use the graph to find the figures relating to his own particular problem, figures which will be more enlightening to his Committee than any statement about correlations.

SELECTION OF PUPILS FIT FOR SENIOR SECONDARY EDUCATION

It is only in regard to senior secondary education that the problems are likely to arise in Scotland. The treatment, however, is quite general and can be applied to any similar type of selection.

The essential difference between this selection and the preceding is that the pass-mark is not fixed for us by the number of scholarships or the accommodation of the schools: we have the responsibility of fixing it so that we shall, as far as possible, admit those who will profit by a senior secondary course and exclude those who would not.

For each size of correlation there is a definite pass-mark at which the number of misfits is a minimum. This, as we shall show presently, can be demonstrated by a general proof, but for the sake of the non-mathematical reader we use a simpler approach.

For the purpose of our explanation we refer to fig. 26.[1] Let us suppose that 16 per cent. of the applicants are unfit for secondary education. The volume CAD will then be 16 per cent. of the total volume. If the pass-mark were fixed at AB, the number of

[1] P. 122.

admit-fails would be the volume CEA, and the number of reject-successes would be the volume BED. The sum of these two volumes would be the number of misfits.

If we know where AB is fixed we have all the data necessary for the calculation of these volumes. We can, therefore, for any size of correlation, make experiments to ascertain the effect of raising or lowering the pass-mark on the number and the composition of the misfits. Table XLI is an example.

TABLE XLI

Effect of changing the pass-mark on the percentage of misfits for normal correlation

Pass-mark at	Failed	Admit-fail	Reject-success	Total misfits
Mean *	50·0	3·1	37·3	40·4
$-\cdot5\sigma$	30·9	6·1	21·1	27·2
$-1\cdot0$	15·9	9·6	9·6	19·2
$-1\cdot5$	6·7	12·6	3·4	16·1
$-1\cdot8$	3·6	13·9	1·6	15·6
$-1\cdot9$	2·9	14·25	1·25	15·50
$-2\cdot0$	2·3	14·54	·95	15·48
$-2\cdot1$	1·8	14·79	·71	15·50
$-2\cdot6$	·5	15·5	·1	15·7
$-\infty$	0·0	15·9	0·0	15·9

$r=+\cdot5$ Number of unfits = 16 per cent.

* In this chapter σ denotes the standard deviation of the Qualifying scores of the applicants for admission to senior secondary courses.

A good many of the principles of selection are concentrated in this Table. When we cut off 16 per cent. of a normal distribution the cut is made at approximately 1σ from the mean. OF $=h$ is, therefore, in this case -1σ. If we draw the pass-mark AB also at -1σ, AB and CD would meet on the major axis of the ellipse, and the position would be the same as for the award of scholarships. The number of admit-fails should be the same as the number of reject-successes, and this is so in the Table.

Studying the figures in the last column we see that the percentage of misfits is a minimum when the pass-mark is drawn at -2σ. If it is drawn above that point the number rises steeply and considerably; below that point the number again rises, but much less steeply.

It is interesting to find that the pass-mark for minimum misfits is so much lower than the line separating successes from failures. The latter is 1σ below the mean, and cuts off 16 per cent. of the pupils; the former is 2σ below the mean, and this would cut off less than 3 per cent.

An important practical point is involved here. Even if we knew the number of fit applicants this would not enable us to fix the pass-mark. If we fixed it in such a way as to admit the proportion we knew to be fit, it would be much too high, and the number of misfits would be well above the minimum. In certain cases, if we wish the misfits to be a minimum we must admit far more than the number who are actually fit.

We note, too, that in this particular case the number of admit-fails at the minimum greatly exceeds the number of reject-successes.

At the pass-mark at which the number of misfits is a minimum, a pupil has a 50 per cent. chance of success in the secondary course; below this pass-mark the pupil has more than a 50 per cent. chance of being a failure. This can be demonstrated in a very simple way from fig. 31.[1]

Let CD represent the line dividing successes in the secondary course from failures, and let RS be the line joining the two horizontal tangents to the ellipse. RS passes through the centre of the ellipse, and on it lie the mid-points of all horizontal chords. Consider now the point E where RS cuts CD. If we draw through E a horizontal chord AEB, AE will be equal to EB, and the volume of the solid on a narrow strip cut at AB will be divided into two equal parts at E. This means that, at AB, half the pupils are successes in the secondary course and half are failures. That is to say, a pupil with the entrance score corresponding to AB has just an even chance of success in the secondary course. With this pass-mark the number of misfits will be the sum of the volume CEA and BED.

If we raise AB slightly (E being the point at which AB cuts CD), we shall reduce the volume CEA and increase the volume BED. But since, in the raised position, AE will be shorter than EB, the amount by which we increase the reject-successes will be greater than the amount by which we reduce the admit-fails. Therefore, by raising the pass-mark above AB we increase the number of misfits. Similarly, we can show that if we lower AB below its present position we again increase the number of misfits. AB is

[1] P. 132.

thus the point at which the misfits are a minimum, and it is the point at which the pupil has just an even chance of success.[1]

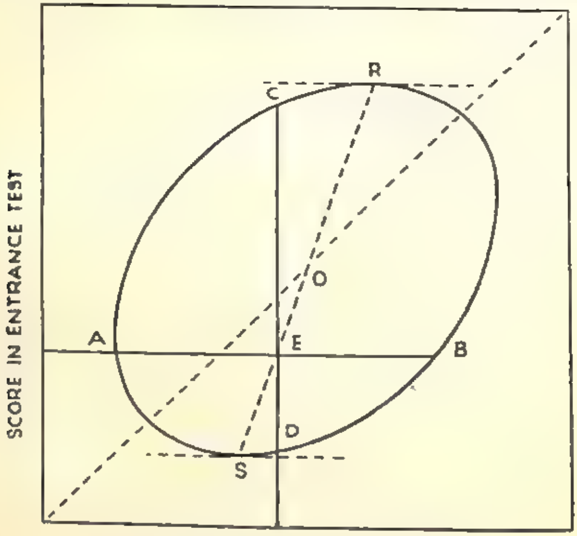

FIG. 31.—Determination of correct pass-mark.

We shall as before call this point at which the number of misfits is a minimum the correct pass-mark.

Curves of Misfits

If we know the correlation between an entrance test and success, we can, for any proportion of unfits in the group of candidates, calculate the number of misfits for any pass-mark. Thus we can plot a graph which will show in a clearer way the effect of raising or lowering the pass-mark.

We can also do this for different sizes of correlation, and see how the curves compare with one another.

Such curves are given in fig. 32, a diagram which should be studied by all who have to deal with selection. It is based on a proportion of fits to unfits which is very close to that found in the

[1] The minimum is always at the point where the regression line of the x-arrays cuts the line dividing successes from failures. For the mathematician, who has read the above popular demonstration with impatience, it may be added that a proper mathematical proof is easy. We take the expression for the sum of the volume CEA and BED as given in footnote p. 123, differentiate and equate to zero. Mr A. S. Robertson, who carried out the operation, found it to be much less difficult than might be expected.

City in which the Inquiry was conducted, and is likely to be reasonably similar to that in other parts of Scotland. We take it that about 72·6 per cent. of the applicants are fit for secondary education and 27·4 per cent. are unfit. Curves for four sizes of correlation are shown.

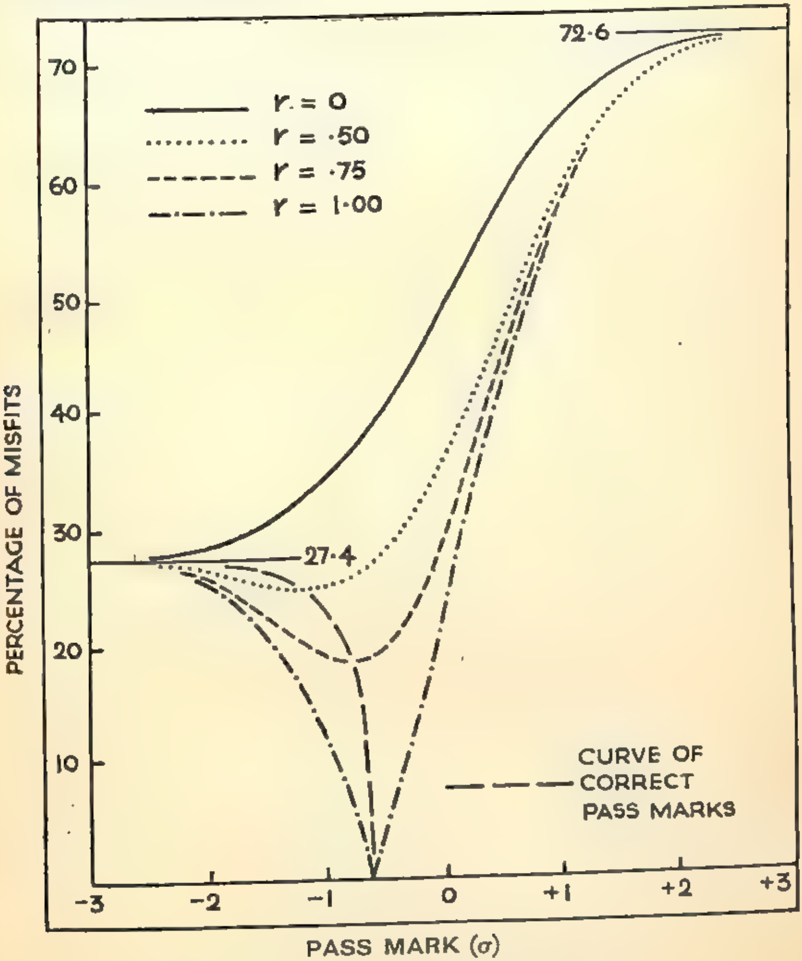

Fig. 32.—Curves of misfits for different sizes of correlation. (Proportion of unfit less than half.)

If the pass-mark is raised above the correct point the number of misfits rises to a maximum equal to the number of fit applicants: if lowered below the correct point, to a maximum equal to the number who are unfit. This is obvious with or without the help of fig. 32.

If the number of fit applicants is greater than the number of unfit, the number of misfits rises more quickly when the pass-mark is raised above the correct point than when it is lowered below it. This principle is clear from a study of the curves in fig. 32, and it will be seen that the effect is more pronounced with the lower correlations. It is an important one, for it shows that the danger, in Scottish

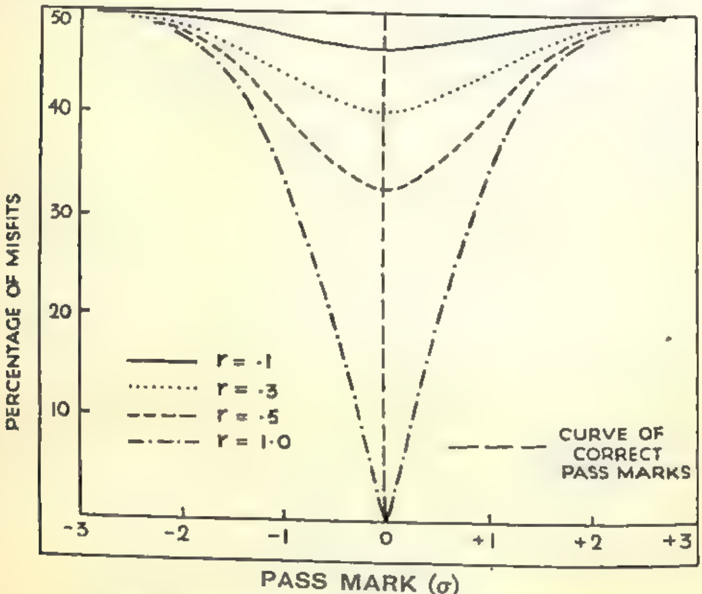

FIG. 33.—Curves of misfits for different sizes of correlation. (Proportion of unfit one-half.)

conditions, lies on the side of a high pass-mark rather than on the side of a low one.

We here come upon a curious and somewhat annoying fact. Nearly all the laws depend upon the point where the vertical CD in fig. 31 cuts the distribution, that is, upon the proportion of fit to unfit candidates. Generally speaking, if a certain law holds when the proportion of unfit is less than half, the law is reversed if the proportion of unfit is more than half. The present principle is a good example. Fig. 32 gives the position when the proportion of unfit is less than half. Fig. 33 shows the situation when half are fit and half are unfit.

The curves now become symmetrical, and the number of misfits rises equally steeply on both sides of the correct pass-marks.

If now the percentage of fits were less than half, say 27·4 per

cent., the curves would be the mirror images of those in fig. 32, and the misfits would rise more steeply on the side of lowering the pass-mark.

This brings out the important fact that, in considering the problems which arise in selective examinations, such as the fixing of pass-marks, the Education Committee must adapt its procedure to the percentage of unfits in the group of candidates. There is, of course, no force in the objection that this could not be known unless all the candidates went on to the secondary course; and that, even then, the information would be available three years too late. All that is needed is a general idea of the proportion, and every Director of Education is in a position to know this. In Scotland, for example, the proportion of unfit candidates for secondary education will almost certainly always be less than 50 per cent. In the City in which the Inquiry was made it is about 29 per cent., and varies little from year to year. Fig. 32 is therefore the case that is relevant to Scottish conditions.

The practical value of an entrance examination depends not merely upon its correlation with later success but also upon the percentage of unfit candidates. By practical value we mean the value to the Education Committee in ensuring that the number of misfits is a minimum.

The truth of the principle can be seen by a comparison of figs. 32 and 33, from which we see that an examination giving a correlation of ·5 with success might be of some use if half the candidates were unfit, but that it would be almost useless if only 27 per cent. were unfit.

If the percentage of unfit candidates is under 30 it is probable that many of the present entrance examinations are a waste of time and money. The principle is stated in a provocative way, but the following are the facts upon which it is based. We have already pointed out that the average correlation between entrance examination and success recorded in four published investigations was ·34; and, judging from the data available, we doubt whether most entrance examinations have a correlation as high as ·4.

If that is not so, our principle has to be modified. But, taking first an examination with a correlation of ·5, we see from fig. 32 that the minimum misfits, if we used the correct pass-mark, would be 25 per cent. If we used no examination at all, and

admitted all the candidates, the number would be 27·43 per cent.: the saving of misfits would thus not be very great, even if we managed to hit the exact pass-mark. But if, as in our experience very often happens, the pass-mark were too high, say $-\cdot 5\sigma$ instead of $-1\cdot 2\sigma$, the number of misfits would be 28·34 per cent.; that is, more than we should have if we admitted all the candidates.

If the correlation were $+\cdot 4$, the minimum misfits with the examination would be 26·55 per cent.; with a correlation of $+\cdot 3$ it would be 27·23 per cent.[1]

This does not mean that an Education Committee should give up its entrance examination—merely that it should take steps to find out the correlation which the examination gives with success, and the normal percentage of unfits among its candidates. If the results, considered in the light of what has been said above, show that the examination is nearly valueless, it should face the facts and mend its ways. This can be done by improving the setting, administration and correction of the examination, and by adopting some of our later recommendations as to the inclusion of other measures in the battery.

We leave as an exercise to the reader the consideration of the case where the percentage of *fit* candidates is 30 per cent. or less and the correlation is about ·4, hoping that he will not conclude that the best plan would be to reject the lot—thus saving not only the cost of the examination but also that of the secondary education.

If the number of fit candidates is greater than the number of unfit, then the higher the correlation between the entrance examination and success the higher is the correct pass-mark. This important result is clear from fig. 32. The point at which we get the minimum of misfits is lower for the lower correlations, the law of fall being indicated by the broken curve. The difference is not very great for correlations between ·75 and 1·0, but below ·5 the correct pass-mark falls very rapidly.

As usual we find the neutral point when 50 per cent. are unfit. The correct pass-mark is then always the same, as can be seen from fig. 33. If the percentage of unfit pupils is more than 50

[1] In Scotland the distribution of candidates will usually be tailed, like that of our senior secondary group (see fig. 57, p. 203). With such a distribution the entrance examination would be of somewhat greater value, the minimum misfits for correlations of $+\cdot 4$ and $+\cdot 3$ being about 25·5 per cent. and 27·0 per cent. respectively.

the law is reversed, and the higher the correlation the lower should be the pass-mark.

In fig. 34 we give the exact curve of fall for the case where 27·4 per cent. of the candidates are unfit, and from this a Director of Education can get a rough idea as to where his pass-mark should be drawn if he can estimate the size of the correlation between entrance test and success.

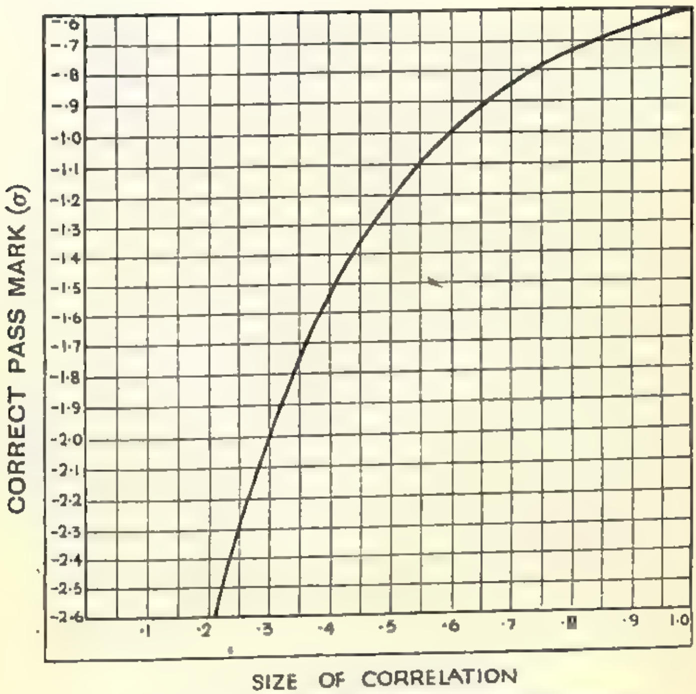

FIG. 34.—Correct pass-mark for different sizes of correlation. (Percentage of unfit 27·4.)

If the number of fit candidates is greater than the number of unfit, the pass-mark should never be fixed above the point at which it will cut off a number equal to the number of unfit. It can easily be demonstrated that if the line separating the fits from the unfits is at $-p\sigma$, then the pass-mark will be highest when $r = 1{\cdot}0$, and it will then be at $-p\sigma$.

The rule may serve as a rough guide to a Director of Education. If he estimates, from the experience of years, that the percentage of unfit candidates is about 27, he knows that the pass-mark

should not be above $-\cdot 6\sigma$. All he has to do, therefore, is to find the standard deviation of his entrance examination marks, multiply it by ·6, and subtract from the mean. The pass-mark should certainly not be above this point, unless the distribution differs considerably from the normal form.

The higher the correlation between entrance examination and success, the more critical is the pass-mark and the more important it is that it should be correctly fixed. The importance of the correct fixing of the pass-mark, in this type of selection, has perhaps not yet been sufficiently realised; and in this principle we draw attention to the fact that the better the entrance test the greater is the danger. This is at once clear from the curves in figs. 32 and 33. If we take an allowable error in fixing the pass-mark to be that giving an increase in the misfits equal to 1 per cent. of the candidates, then with a correlation of 1·0 our latitude in drawing the pass-line in fig. 32 would be only $\pm \cdot 05\sigma$. With a correlation of ·75 it would be $\pm \cdot 45\sigma$.

Even with a perfect test (that is, giving a correlation of 1·0 with success) we could have a large number of misfits if the pass-mark were incorrectly fixed. This is not an independent principle, but its importance is sufficient to warrant separate statement. In striving to improve our selection it is not sufficient to perfect the tests. As we have seen above, a very good test is in some ways a positive danger.

We shall later make some suggestions as to ways in which the dangers may be avoided; but in the meantime we give two numerical illustrations of their magnitude:

If, with a test giving a correlation of 1·0, we fix the pass-mark 8 per cent. too high (assuming an ordinary distribution of examination marks), we should have the same number of misfits as with a test giving a correlation of ·75, provided that the correct pass-mark were used.

If, in our own senior secondary group of 416, with our best battery (giving a correlation of ·8), we had fixed the pass-mark 10 per cent. too high, we should have increased the misfits by 28. This is equivalent to using a test with a correlation of $+\cdot 57$ and the correct pass-mark.

If the number of fit candidates is greater than the number of unfit, then at the correct pass-mark the number of admit-fails

will be greater than the number of reject-successes. The purport of this principle is that, if an Education Committee adopts our suggestion as to the correct pass-mark, it is not erring on the side of refusing to give the child his chance.

The further the percentage of unfit candidates falls below 50 the greater will be the preponderance of admit-fails over reject-successes at the correct pass-mark. The size of this effect, for a correlation of +·50, is shown below:

Percentage of unfit candidates	Percentage at the correct pass-mark		
	Admit-fails	Reject-successes	Misfits
50·0	16·7	16·7	33·3
30·8	21·1	6·1	27·2
15·9	14·5	·9	15·5
11·5	11·0	·3	11·4

If the number of fit candidates is greater than the number of unfit, the lower the correlation between the entrance test and success, the greater is the preponderance of admit-fails over reject-successes at the correct pass-mark. In fig. 35 we have plotted the curves of admit-fails and reject-successes at the correct pass-marks for different sizes of correlation. The curves are in respect of the case where 27·4 per cent. of the candidates are unfit, which is nearly the normal proportion in the City in which the Inquiry was made.

This shows that, under Scottish conditions, the poorer the entrance test the more we err on the side of giving the pupil his chance, provided we use the correct pass-mark.

If the number of fit candidates is greater than the number of unfit, the lower the correlation between the entrance test and success the greater the number of pupils we have to admit. This is an important practical law, for it has financial implications. We assume, of course, that the Education Committee desires to have the misfits a minimum; and, if so, it is clear that it is bad economy to save money on the entrance testing.

Taking what we have regarded as a fairly normal percentage of unfit candidates for Scotland, that is 27·4 per cent., the percentage of candidates that we should have to admit for different sizes of correlation between entrance test and success is as follows:

Correlation	Percentage to be admitted
1·0	72·6
·9	74·9
·8	77·3
·7	80·5
·6	84·1
·5	88·5
·4	93·3
·3	97·7
·2	99·9
·1	100·0
0·0	100·0

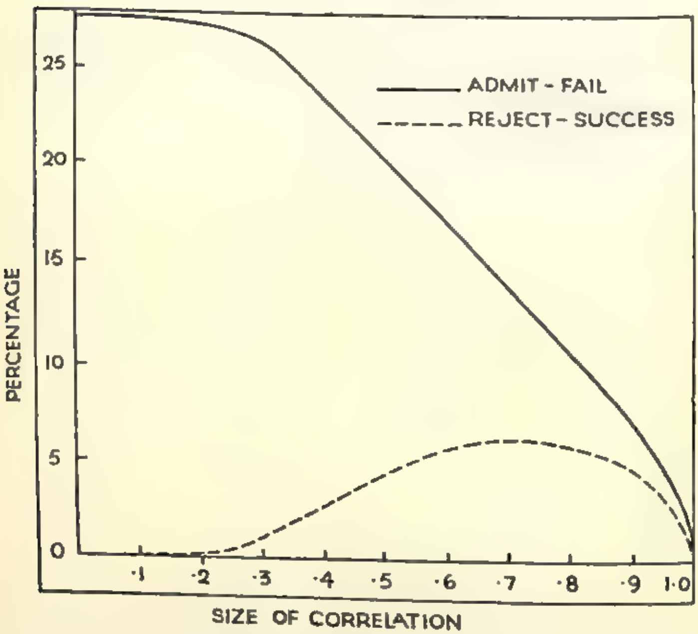

FIG. 35.—Curves showing relative proportions of admit-fails and reject-successes at the correct pass-mark for different sizes of correlation. (Percentage of unfit 27·4).

From this it is clear that, if we are to deal fairly with the pupils and make the most economical use of the national resources of money and talent, we must admit to the senior secondary schools more pupils than the number who are fit. If we had a perfect test and used the correct pass-mark, we should admit 72·6 per cent.; with our own best battery ($r = \cdot 8$), we should have to admit 77·3 per cent. With an examination, giving a correlation of about ·4, we should have to admit 93·3 per cent. If there were

1,000 candidates this would mean that we should have to provide secondary education for 160 more pupils, a large proportion of whom would be failures. And it will be clear by now that an Education Committee cannot counteract the effects of a poor examination by using a higher pass-mark than the correct one for that examination.

If about 27 per cent. of the candidates are unfit, and if the number of failures in the secondary course is less than 9 per cent. of the number admitted, the system of selection is unsound. This is merely a rough rule for Scottish conditions, and it draws attention to an erroneous opinion that is sometimes held. Some Education Committees, in experimenting with different systems of selection, collect statistics as to the success or failure of the pupils admitted to the senior secondary schools; and in discriminating between such systems it is occasionally assumed that the one producing the fewest failures in the secondary school is the best. This is far from the truth. The fact that system A admits fewer failures than system B probably means that A is the less sound system; but the position is that we cannot judge on this basis alone. We must take into account the correlation, the proportion of unfit candidates and the position of the pass-mark.

We have already established that it is vain to aspire to a system with no misfits. The question then arises as to what should be the percentage of failures among the pupils admitted if the system is as sound as present conditions permit. Our answer depends on the percentage of unfit candidates; but, taking this to be approximately 27 per cent., and assuming normal correlation, there should be 14 per cent. of failures among those admitted. If the distribution had a tail of weak pupils, as in our Inquiry group, 11·5 per cent.[1] of those admitted should turn out failures. These figures are based on our best battery; and since the minimum number of admit-fails would be greater with a poorer one, it will be seen that our figure of 9 per cent. is conservative.

Curve showing Predictive Significance of Correlations in the Selection of pupils for Secondary Education

The nature of the curve depends on the proportion of unfit candidates; but if we take this to be 27·4 per cent. of the total, we obtain a result applicable with reasonable accuracy

[1] If we used the additional principles mentioned in Chapter XIV the figure would be 10·1 per cent. See also Chapter XVII, p. 173.

to Scottish conditions. Fig. 36 gives the *minimum* number of misfits for each size of correlation as a percentage of the minimum number for zero correlation; that is, it assumes that the correct pass-mark is always used.

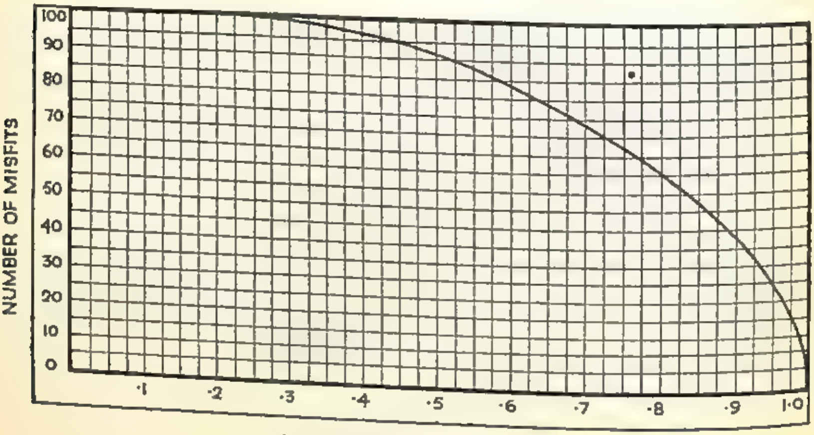

FIG. 36.—Significance of different sizes of correlation in selection of pupils fit for secondary education.

NOTE.—The *general laws* stated in this chapter will hold for all distributions likely to be met in practice, but *numbers and percentages* based on the assumption of normality should be used only as rough guides. Fig. 36, for example, gives the numbers of misfits for different sizes of correlation if the distribution of candidates is normal. We think it probable that, in Scotland, the distributions will usually have the tailed form of our senior secondary group, and, if so, the numbers would be as follows:

Correlation	Minimum number of misfits
0·0	100
·1	100
·2	100
·3	99
·4	96
·5	87
·6	75
·7	61
·8	51
·9	35
1·0	0

CHAPTER XV

ATTEMPT TO IMPROVE SELECTION BY SPECIAL CONSIDERATION OF BORDER-LINE CASES

So far we have assumed a hard-and-fast pass-line. We now consider whether a careful scrutiny of border-line cases might justify the admission of certain pupils just below the pass-mark or the rejection of some just above it, thus improving our selection.

One advantage of the misfits technique is that we can make experiments on such problems with an ease and speed much appreciated after the laborious correlation procedure.

Initial Standard Number of Misfits

For the purpose of the experiments which follow we use the complete battery IQ + Q + S + Ts, for which the minimum misfits in the senior secondary group are as follows:

Number in group	Admit-fails	Reject-successes	Misfits
452	39	20	59

These, then, are our standard figures in relation to which we assess the value of any proposed principle.

Supplementary Principles involving Use of Test and Examination Scores

imposing a minimum in each subject for all pupils

The subjects are intelligence, English and Arithmetic. English is taken to be Qe + E + Tes, and Arithmetic Qa + A + Tas. By imposing a minimum of 0 sigma we mean that a pupil is rejected if he falls below 0 sigma in any subject, no matter how high his average score may be.

We give below the numbers of misfits with different minima:

Minimum	Admit-fails	Reject-successes	Misfits
+ ·6 sigma	20	62	82
+ ·4	26	33	59
+ ·2	32	26	58
0·0	34	22	56
− ·2	38	21	59
− ·4	39	20	59
− ·6	39	20	59
Standard figures	39	20	59

In interpreting this and the following figures we have to bear in mind the difficulty that, while the complete group is reasonably large, the numbers affected by a principle may be small. There is no satisfactory escape from this; but we can introduce some of the steadiness which comes from the numbers above and below by using graphs. By examination of the sweep of these, after smoothing, we can usually tell whether or not a critical figure is likely to be accidental. In the above case when we plot misfits against the minima we find confirmation that there is likely to be a minimum at 0 sigma, but not quite so low as 56.

With this general reservation we hazard the following conclusions:

A small reduction in the number of misfits can be made by imposing a minimum in the separate subjects.

The minimum has to be very exactly determined: if it is too low it gives no improvement; if it is too high the number of misfits is greatly increased.

The practical import is that the principle should be used only if the minimum is scientifically determined by a method which we shall explain later.

IMPOSING A MINIMUM IN EACH SUBJECT FOR WEAKER PUPILS ONLY

To impose the minimum only on pupils with average scores of +1·0 sigma and under is the same as to demand a higher pass-mark, namely, +1·0 sigma, if the pupil falls below the

minimum in any part of the examination. The numbers of misfits with different minima are shown below:

Minimum	On pupils below	Admit-fails	Reject-successes	Misfits
0·0 sigma	+1·5 sigma	34	22	56
	+1·0	35	22	57
	+ ·9	37	22	59
	+ ·8	37	21	58
	On all pupils	34	22	56
+ ·2 sigma	+1·0	34	24	58
	+ ·8	36	22	58
	On all pupils	32	26	58
+ ·3 sigma	+1·0	31	26	57
	+ ·8	34	23	57
	On all pupils	31	28	59

Clearly nothing is gained by limiting the application of the minimum to the weaker pupils.

IMPOSING A MINIMUM (*a*) IN IQ ONLY, (*b*) IN ENGLISH AND ARITHMETIC SEPARATELY BUT NOT IN IQ

We found that both of these gave poorer results than the imposition of a minimum in each subject as in the preceding sections.

ALLOWING A LOWER PASS-MARK IF THE PUPIL HAS HIGH ABILITY IN ONE SUBJECT

This is the converse type of allowance. It is based on the idea that a pupil who falls just below the pass-mark but has a very high IQ or very high ability in English or Arithmetic, may have a better chance of success in the secondary course than a pupil who is just above the border-line in all subjects. The number of misfits when a lower pass-mark is allowed for pupils with high ability in any subject is as follows:

Test of high ability	Lower pass-mark allowed	Admit-fails	Reject-successes	Misfits
Over +1·3 sigma	+·6 sigma	39	19	58
	+·5	40	19	59
	+·4	19	41	60
	+·3	19	41	60
Over +1·1 sigma	+·6	19	42	61
	+·5	18	43	61
	+·4	18	45	63
	+·3	18	45	63
Standard figures		39	20	59

It is therefore most unlikely that any great improvement could be made in this way. Our results show a slight reduction if we allow a pass-mark ·1 sigma lower for pupils who are above +1·3 sigma in any one subject.

APPLYING BOTH (*a*) A MINIMUM IN ANY SUBJECT, AND (*b*) A LOWER PASS-MARK IF THE PUPIL HAS HIGH ABILITY IN ONE SUBJECT

These supplementary principles may conflict when combined. Thus a pupil whom we have passed on account of high ability in English may fail because he is below the minimum in Arithmetic. We have therefore to fix a definite order of application, and we found the following to be the best:

(1) Reject all pupils who fall below 0 sigma in one subject.
(2) Admit pupils between +·6 sigma and +·7 sigma if they are above +·1·3 sigma in any subject, and are not rejected by reason of (1).

By this procedure we find that the following are the numbers of misfits:

Admit-fails	Reject-successes	Misfits
34	21	55

The resulting improvement (4 misfits out of 59) is considerable; and the above now become our standard figures for the application of further principles.

PROTECTING THE PUPIL AGAINST A FLOP IN ONE TEST OR EXAMINATION

We have given an example of this in fig. 19, where it is clear that some accidental disaster has overtaken the pupil in the Arithmetic examination. In diagnosing a flop it is desirable that we should have at least two other reliable measures *in the same subject*; and the drop should be greater than those that could be explained by normal fluctuations of performance and marking.

We cannot apply this principle to a low IQ as we have no other tests in the same subject to act as a standard. Yet with many pupils we had the strongest suspicion that the IQ result was of the flop type, and that this increased the number of misfits. We therefore make two suggestions. It is desirable that further research be undertaken towards obtaining comparable IQs from different group tests. It would then be an advantage if Education Committees were to arrange for the application of a group or individual intelligence test to all primary pupils at least a year before the Qualifying stage. The results of such testing could be used to protect the child against the danger to which we have referred, and could be put to many other uses. Another method would be to apply an individual test in all suspected cases.

After some experiment it was found that the best system was to count a score a flop if:

(1) The other two scores in the same subject were within ·3 sigma of each other.

(2) The third score was at least ·7 sigma below the lower of the other two.

The effect of protecting against a flop in one test is shown below:

	Admit-fails	Reject-successes	Misfits
After allowance for flops	34	19	53
Standard figures . .	34	21	55

The principle thus makes a reduction of 2 misfits in 55; and, while the numbers are so small that we cannot vouch for the result with any great degree of statistical assurance, our experience of several years of practical selection warrants the recommendation that it should be accepted.

It will be understood that the number of flops in our group was much greater than 2. The difficulty, indeed, with all these

subsidiary principles is that they are two-edged weapons. A number of pupils who were satisfactorily housed with the reject-fail group, and whom we should have been only too glad to leave there, were transferred by our flop principle to the admit-fails. They thus *increased* the number of misfits.

Supplementary Principles involving Use of Other Data

We now consider whether we could improve our selection by going beyond the test and examination results and taking into account such factors as health, industry, home conditions and personal qualities.

ALLOWANCE FOR HEALTH

It is unlikely that anyone, outside ultra-scientific circles, would think of including health ratings in the scores before the pass-mark is determined. The most that we could do would be to take it into account in extreme gradings or border-line cases, and one method would be to reject all who have a very low health grading, irrespective of what their average score may be. To test this we tried rejecting all with a health grading of E, and found that we increased the misfits by 1; when we rejected all with D or E the increase was 15.

Another method would be to take health into account in border-line cases only. If a pupil were just below the pass-mark but had a health grading of A, there would be a case for admitting him in preference to one who was just above the line but had a health grading of E. To test this idea we used a principle similar to that adopted in allotting age-allowances. We gave a bonus of ·2 sigma for an A, ·1 sigma for a B, a reduction of ·1 sigma for a D, of ·2 sigma for an E. This amounts to allowing a slightly higher or lower pass-mark according to the health grading. The effect of different allowances for health on the number of misfits is shown below:

Health allowance	Admit-fails	Reject-successes	Misfits
A + ·4 : B + ·2, etc. .	43	19	62
A + ·2 : B + ·1, etc. .	37	19	56
A + ·1 : B + ·05, etc. .	37	19	56
A + ·02 : B + ·01, etc. .	35	19	54
Standard figures .	34	19	53

The conclusion is only too obvious, and is confirmed by the method mentioned on p. 144. If we plot number of misfits against size of health allowance the curve reaches the minimum when the health allowance is zero.

ALLOWANCE FOR INDUSTRY

Fig. 37 will illustrate the temptations to which one is exposed in dealing with these Qualifying problems. The correct pass-mark

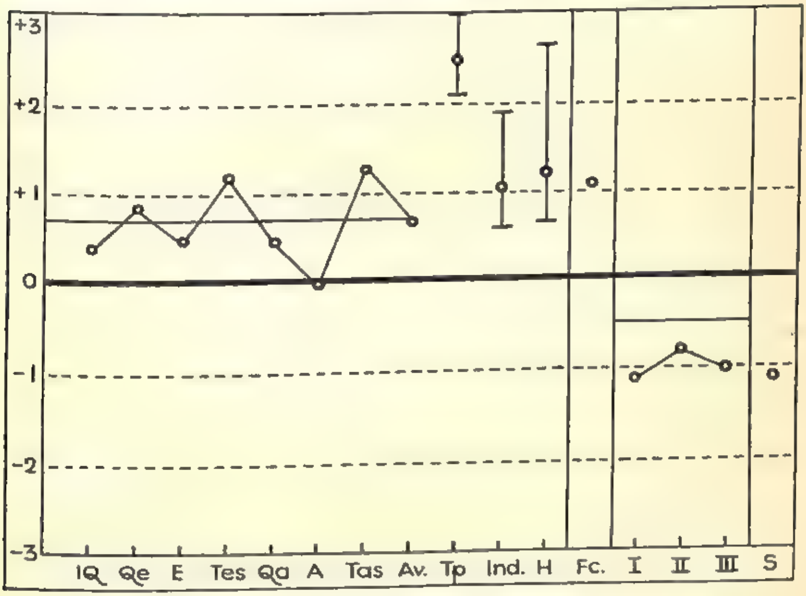

FIG. 37.—Case where allowance for industry, etc., would lead to error.

is +·70, and this girl's average score is +·69; yet she has very pronounced practical ability; is well above the average in industry and in health; has originality and special talents in Art and Music; comes from a good home; and her primary head master is confident that she will be a success. Surely we were intended to admit this girl and give her the chance of a senior secondary education which may be the road to a career in music or art. We should in this case have been wrong.

In fig. 38 we give the opposite type of case, where we should have been saved from error by taking into account the teacher's forecast and industry grading.

As with health, we first rejected all pupils with an E in industry. This made no change in the number or composition of the misfits. Rejection of all with D or E ratings left the total number of misfits unchanged, but slightly lowered the number of admit-fails relative to the reject-successes. We also tried the method of assigning a bonus or penalty in accordance with the grading,

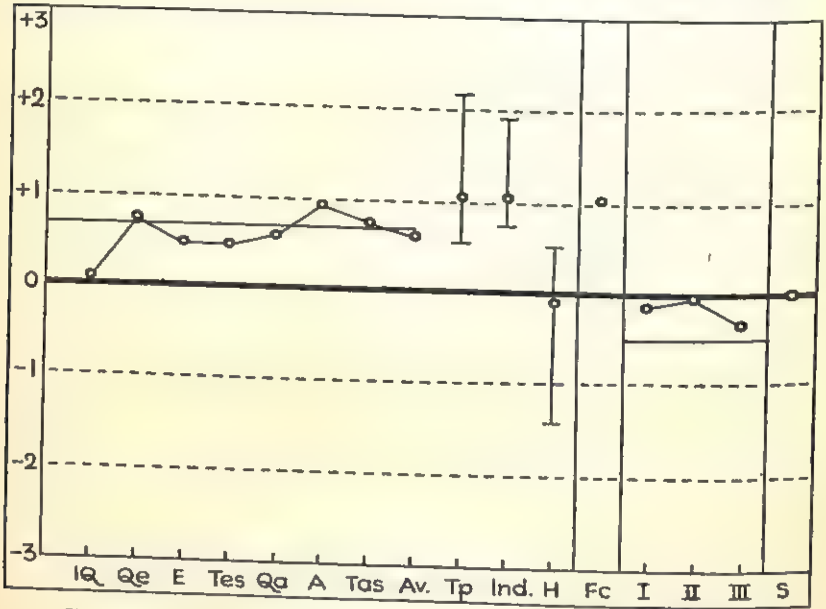

FIG. 38.—Case where allowance for industry, etc., would improve selection.

thus taking industry into account only in border-line cases. The effect of such allowances for industry on the number of misfits is shown below:

Industry allowance	Admit-fails	Reject-successes	Misfits
A + 1·0 : B + ·5, etc.	49	12	61
A + ·8 : B + ·4, etc.	45	12	57
A + ·6 : B + ·3, etc.	43	14	57
A + ·4 : B + ·2, etc.	41	14	55
A + ·2 : B + ·1, etc.	40	16	56
A + ·1 : B + ·05, etc.	39	18	57
A + ·02 : B + ·01, etc.	36	19	55
Standard figures	34	19	53

We can only conclude that nothing is to be gained, even in border-line cases, by taking account of industry gradings as obtained in our Inquiry. Perhaps one reason for this is that the pupil's industry will be reflected in his test and examination results and is thus already taken into account.

ALLOWANCE FOR TEACHERS' FORECASTS

On the rating cards the primary teachers gave forecasts of the pupils' success in each of the four main types of post-primary

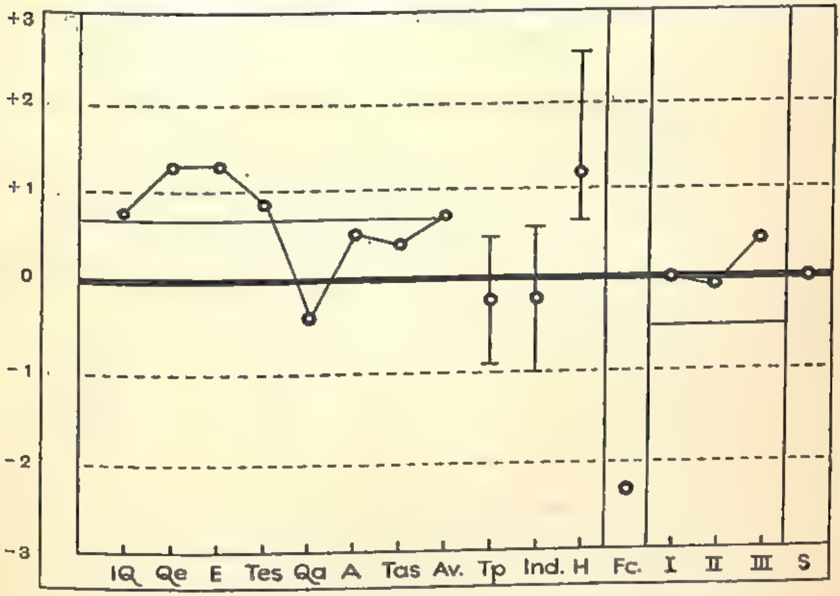

FIG. 39.—Case where allowance for teacher's forecast would lead to error.

course. A five-point scale was used, in which A indicated distinguished success, B clear success, C doubtful success, and so on.

We had high hopes of reducing the misfits through their use, for we noted frequently that pupils whose success was greater than we should have expected from their test results had forecasts of A or B. But once again we had a compensating number of cases, like that shown in fig. 39, where the forecast was clearly wrong.

This pupil has an average score of +·72 sigma and therefore falls just above the pass-mark. But her primary teacher reports that she is untidy, slow in speech, movement and learning, and grades her as a hopeless failure in a senior secondary course. Surely, on such additional information we should bring her below the line. The right-hand side of the profile shows that, had we done so, we should have made a decided blunder. The pupil took a one-language course, attained nearly the average of the secondary group in her first two years, and showed signs of improvement even on that satisfactory standing.

By failing all pupils with a teacher's forecast of E we increased the misfits by 2, while rejection of all with forecasts of D or E gave an increase of 4.

Using the forecasts in border-line cases only we get the following results:

Forecast allowance	Admit-fails	Reject-successes	Misfits
A+1·0 : B+·5, etc.	41	18	59
A+ ·8 : B+·4, etc.	40	22	62
A+ ·6 : B+·3, etc.	40	21	61
A+ ·4 : B+·2, etc.	39	20	59
A+ ·2 : B+·1, etc.	39	20	59
A+ ·1 : B+·05, etc.	37	21	58
A+ ·02 : B+·01, etc.	36	20	56
Standard figures	34	19	53

The conclusion is that the use of the teachers' forecasts even in border-line cases increases the number of misfits.

ALLOWANCE FOR PERSONAL QUALITIES

Primary teachers were asked to note any marked personal qualities of the pupils. This was not a very scientific way of collecting the data, but we felt that it would be sufficient for a preliminary exploration of the problem.

The view has often been expressed that existing systems of selection are based too exclusively upon intellect and scholastic attainment, and that success in higher education may be dependent not merely upon these but also upon certain personal qualities. At first sight our case histories seemed to bear this out. In reviewing the profiles we encountered a number where a pupil just under the admission line had a number of good qualities and

proved a success; and conversely where a pupil just above it had marked bad qualities and was a failure. The matter was therefore worthy of further study.

To reduce the results to a form amenable to statistical treatment was far from easy, but the following rough procedure was finally adopted. We divided the qualities into three categories, good, bad and neutral. In so doing we were guided largely by common sense, and partly by the results of previous studies, which suggest that the most important qualities from this point of view are those like logical and critical ability, independence of thought, eagerness to learn, and are not to be found on the side of emotional life and character.[1] In the good category we included such qualities as perseverance, reliability, earnestness, ambition; in the bad group, carelessness, indifference, laziness, ready discouragement, and so on. To the neutral category we relegated all those that seemed to be irrelevant to success in the secondary course, for example, amiability, distinct charm of manner, good nature, sensitiveness, timidity. In order to take the qualities into account in border-line cases we gave a bonus for each good quality and a penalty for each bad one. The effect of these allowances on the number of misfits is shown below:

Bonus or penalty for personal qualities	Admit-fails	Reject-successes	Misfits
·3 sigma	39	18	57
·2	38	20	58
·1	39	19	58
·05	37	18	55
Standard figures	34	19	53

Once again we get the disappointing result that, however well the idea may work in individual cases, it is not a success when generally applied. If, therefore, personal qualities are to be used in selection, the data must be obtained by a better method than that used by us.

There is, however, a reason why the use of personal qualities in selection is unlikely to make any great improvement; as with industry, the qualities which matter are already reflected in the pupil's test and examination results.

[1] O. Bobertag, *Schülerauslese*. Berlin, 1934. P. 64. Privately printed for the International Examination Inquiry.

ALLOWANCE FOR HOME CONDITIONS

By this time we had abandoned all hope of improvement through the use of such supplementary data, but for the sake of completeness we felt we should not leave this last stone unturned. We divided the reports on home conditions into good, bad and neutral, and examined the effect of taking these into account at the border-line:

Bonus or penalty for home conditions	Admit-fails	Reject-successes	Misfits
·4 sigma	44	15	59
·3	40	17	57
·2	41	18	59
·1	37	18	55
·05	37	19	56
Standard figures	34	19	53

Despite this negative result, which may be due to our method of collecting the data, we feel that home conditions do have an important bearing upon a child's success in the secondary school, and we shall return to the point at a later stage.

ALLOWANCE FOR COMBINATIONS OF THE SUPPLEMENTARY DATA

It is just possible that, while we secure no improvement by taking the additional factors separately, we might do so by taking them together. One pupil might, for instance, survive with an E in industry, another with an E in health, another with extremely bad home conditions. But if the *same* border-line pupil had all three we might be justified in prophesying that he would be a failure.

We tried various combinations without any resulting improvement; and finally we included every rating. For each measure the allowance made was that giving the least unsatisfactory result when it was taken separately, and an age allowance was added. The result was that the total number of misfits was increased from the standard figure of 53 to 61.

MORE SCIENTIFIC SYSTEMS OF MAKING ALLOWANCES

So far we have used the straightforward system of giving a bonus for an A or a B and a penalty for a D or an E. This is probably the only system ever likely to be used in practice, but it has a defect which may be indicated by the following figure:

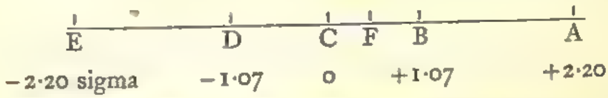

Fig. 40.—The need for a more scientific system of making allowances.

On the assumption that the industry gradings are normally distributed, the means of the categories A, B, C, D and E would be as shown. If, at C, the pupil had a 50 per cent. chance of success, the system which we have used would be sound; but if the point for an even chance of success were at F (that is, if a pupil had to be more industrious than the average to have an even chance of success in the secondary school), a C grading should get a penalty, and the bonus for a B would be less than the penalty for a D. A more scientific system would therefore make the bonuses and penalties proportional to the distances of the means of the categories from F.

Industry

The data determining the point at which the pupil has an even chance of success have some interest in themselves, and are given below:

Industry grading	Successes	Failures	Total	Percentage failures
A	52	2	54	3·70
B	192	29	221	13·12
C	73	76	149	51·01
D	2	20	22	90·91
E	0	5	5	100·00

There is a clear correlation between industry and success; and the point for an even chance comes out to be −·15 sigma. As the mean of C category is −·18 sigma, the straightforward system of bonuses is here justifiable.

Teachers' Forecasts

The percentage of failures in the various categories of the teachers' forecasts are as follows:

Forecast grading	Successes	Failures	Total	Percentage failures
A	52	2	54	3·70
B	214	30	244	12·30
C	41	59	100	59·00
D	8	31	39	79·49
E	2	10	12	83·33

There is again evidence of correlation, but it is interesting to note that nearly 4 per cent. of those whom the teachers expected to be distinguished successes have actually proved to be failures, and that about 17 per cent. of those for whom they predicted hopeless failure have been successful. By the time the reader has finished this Report he will, we trust, fully realise that this is no reflection upon the care or judgment of the teachers.

At this point we interpolate a Table prepared by Mr Robert Robertson, M.A., which throws additional light upon the reliability of teachers' forecasts.

TABLE XLII

Teachers' forecasts of success in senior secondary course for different groups

Group	N	Percentage				
		A	B	C	D	E
Senior secondary . .	452	12·1	54·6	21·8	8·8	2·7
Junior secondary . .	1,930	·6	17·7	34·4	32·9	14·4
Special classes for backward children . . .	509	0	·2	·9	20·0	78·9

These data show that the primary teachers can make a fair sifting of the material at the Qualifying stage, but that we could

not rely solely upon their estimates. If, in the senior secondary group, we had used the teachers' estimates alone, we should have had 42 misfits in the A, B, D and E categories; and, in addition, 100 pupils were graded C.

Finding that the point for an even chance is +·20 sigma, we constructed two scientific bonus systems, in one of which A got a bonus of +·20 sigma, and in the other +·10 sigma. Neither was of any avail, the total number of misfits remaining obstinately the same, though the composition was considerably altered.

Health and Personal Qualities

For these factors the correlation is so small that the point for an even chance of success cannot be determined. The percentage of failures in the various groups of personal qualities is given merely for its general interest to any who may contemplate a more scientific attack on this problem:

Personal qualities	Successes	Failures	Total	Percentage failures
3 good qualities .	24	4	28	14·3
2 good qualities .	61	5	66	7·6
1 good quality . .	79	16	95	16·8
No marked qualities .	136	79	215	36·7
1 bad quality . .	18	26	44	59·1
2 bad qualities . .	2	2	4	50·0

AGE ALLOWANCES

We give some results bearing upon one particular aspect of this much-discussed topic.

Relevance of the Age-Allowance Procedure to Scottish Conditions

The introduction of age allowance usually springs from a desire to be fair to all candidates in the competition for a limited number of scholarships or places in secondary schools. If the examination is confined to pupils between 11 and 12 years, it is felt that the pupil of 12 has an unfair advantage over the pupil of 11. The younger child should therefore be assigned a bonus; and the test of fairness is that the percentage of awards should be uniform throughout the age-range.

Morally the idea is excellent; and it is a pity that the need for it

should arise in this system of limited competition whose educational morality is open to question. Even here a wider view of fairness is probably necessary; and it is certainly so in Scottish conditions at the Qualifying stage. Our bursaries are intended to meet the needs of all pupils desiring a secondary education, and fitted to profit by it, who would be debarred by reason of the expense involved. Our selection system would be contrary to the statutes if it involved the exclusion from secondary education of pupils who had more than an even chance of success. So far as we have competition, it should be limited only by this ability to profit, not by a fixed number of bursaries or places in secondary schools; and it follows that our test of fairness must take the child's success in the secondary course into account. If we admit a young pupil, by giving him a bonus for his youth, our fairness may be costly to him should he prove to be a failure.

Effect of Straightforward Age Allowances on the Number of Misfits

We first tried the effect of the simplest form of age allowance, in which a bonus of so many marks is given to a pupil for each month by which he is below the maximum age. A common allowance of this type is 1·2 per cent. a month, and to assist in interpreting the following data it may be mentioned that ·06 sigma is equivalent to about 1 per cent. The following are the effects of different age allowances on the number and composition of the misfits:

Age allowance per month	Admit-fails	Reject-successes	Misfits
·10 sigma	49	37	86
·05	42	27	69
·04	39	24	63
·03	39	24	63
·02	36	21	57
·01	37	20	57
−·01	39	20	59
−·03	44	22	66
−·05	56	31	87
−·10	74	45	119
Standard figures	34	19	53

To convince the reader that the result is not due to chance fluctuations of small numbers we give the graph, from which it will be seen that the general tendency of the curve is clear. The negative allowances are, of course, merely of the nature of a mathe-

matical diversion. They mean that we give the old pupil a bonus simply for being old, and that we *penalise* the young pupil for his youth. Even with as modest an allowance as $\frac{1}{6}$ per cent. a month we increase the number of misfits, and the higher the allowance the greater is the increase. An allowance of 1 per cent. a month would be almost as harmful as a negative allowance of ·5 per cent.

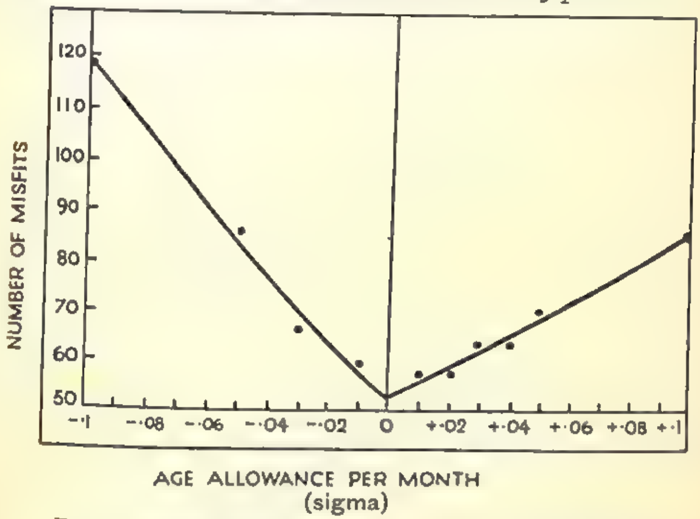

FIG. 41.—Effect of different age allowances on the number of misfits.

Effect of more Scientific Age Allowances on the Number of Misfits

Certain defects of the straightforward type of age allowance are brought out by figures 1, 2 and 3,[1] which show that the age allowance for each month should not be the same for the weak as for the strong pupils. A scientific allowance should therefore depend on the pupil's score.[2]

From our data we prepared the following differential allowance:

Pupil's score	Age allowance per month
Over +1·32 sigma	+·03 sigma
+·50 to +1·31	+·04
−·50 to + ·49	+·05
Below −·50	+·03

[1] Pp. 10–11.

[2] G. H. Thomson, "The Standardisation of Group Tests," *British Journal of Psychology*, vol. ii, Parts II and III. J. B. T. Davies and G. A. Jones, *The Selection of Children for Secondary Education.* London: George G. Harrap & Co., Ltd., 1936. Chapter vii.

With this system the number of misfits is increased from 53 to 66.

These results should be considered in conjunction with those given earlier in regard to correlations with age, the high correlation between intelligence score and intelligence quotient, and the close agreement between the follow-up correlations for these.[1]

Summary

This has been a depressing chapter of negative results and discarded principles; yet it has cleared the air, and others may profit by our mistakes. For the benefit of future investigators we would emphasise one important point. A study of these supplementary factors in border-line cases can be most satisfactorily made if no selection system has been operative. In this we were especially fortunate, for our Inquiry deals with the last group admitted to senior secondary schools without a fairly strict selection.

Despite our disappointments, we have gathered a crop of 6; and a saving of 6 misfits out of 59 is worth making. The correlation of the complete battery with success was ·800; if we use the supplementary principles the result is equivalent to a correlation of ·854.

The only additional principles which we can recommend are:

- The imposition of a minimum for each separate subject.
- The allowance of a slightly lower pass-mark if the pupil shows high ability in one subject.
- The protection of the pupil against the effects of a flop in one subject.

We shall show later how these principles could be used in practice.

The extent of our success in reducing the number of misfits is shown below. The group numbered 452, of whom 320 were fit for secondary education:

Method of selection	Number of misfits
All applicants admitted . . .	132
Complete battery ($r = +\cdot 8$) with pass-mark of $+\cdot 7$ sigma	59
The same, with above additional principles	53

[1] See Note by Professor Godfrey H. Thomson on p. 161.

AGE ALLOWANCES

Note by Professor Godfrey H. Thomson

This fact, that giving age allowances increases the number of misfits in the senior secondary courses, is in agreement with what I too have sometimes found in researches in England, namely that entrance examination marks without age allowances correlated slightly better with secondary school performance than did marks with age allowances. These results might appear to be an argument against age allowances. But I do not draw that conclusion, for *the object of an age allowance is not to improve prediction but to do justice to children born in different months of the year.*

It is true that age allowances, at any rate at first sight, seem to be less imperatively necessary at this stage in Scotland than in England. In England the examination is more competitive, and is almost invariably held only once a year, whereas in some Areas in Scotland it is held every six months. In England a pupil as a rule must sit the examination with his age-group, and save in exceptional circumstances has no further chance; in Scotland he may wait until a later age, and may even then be allowed to sit again if unsuccessful.

But these differences are more apparent than real. The examination in Scotland is really competitive, even if it be less so than in England. There is a mark below which the candidate does not get a bursary; a mark below which he is not recommended to go to a senior secondary school but rather to a junior secondary school; a mark below which he is not allowed to pass on to any kind of secondary school. The difference between England and Scotland is rather that the competition in Scotland is at a different level of ability; but it is still there.

Not by any means in every Area in Scotland is the examination held every six months; and even where it is, one of the examinations usually has many more candidates than the other. And although there may be another examination only six months ahead, a boy who has been recommended to go to a *junior* secondary school may take the advice, and not try again. Indeed he is not always allowed to do so.

I feel strongly then that age allowances should be given, at least over the six or twelve months within which the main body of candidates were born, and a reasonable period beyond that with certain precautions. If this is not done, then the younger children born within that period of six or twelve months will win bursaries, or will get into secondary schools, in smaller numbers than the older ones, and this is obviously unfair. If it be replied that these young candidates have another chance next year, or even in six months' time, I have already answered that in part; and in any case failure is discouraging. But even if they try again, they are still at a disadvantage if no age allowances are given; for among the group of children who are sitting a second time, these are still the youngest.

I admit that it is very difficult in Scotland to calculate the correct age allowance because of the extremely heterogeneous nature, in age, of the group of candidates, and the unduly wide age-range. But these difficulties may be diminished—they are much less in some counties—and are not insuperable.

As long as examinations are held only annually, or in some cases every six months, it *must* be right to give an age allowance to make a child born in June have the same chance as one born in January, even if this apparently increases the number of misfits in the secondary school. Nor is the reason why it does so far to seek: it is probably because the secondary schools do not give an age allowance over the twelve months in judging a child. They agree with the entrance marks without age allowance because they themselves give no age allowance. Both procedures are unfair, and to that extent are in agreement. Age allowances are sometimes, by those opposed to them, called a premium on youth. They are not that. When scientifically applied they are a device to compensate for the unfair premium on age.

CHAPTER XVI

FINAL ORDER OF PREDICTIVE VALUE OF THE VARIOUS TESTS AND BATTERIES

The Evidence

To the original follow-up correlations for senior and junior secondary schools we have now added the following evidence:

- Best weighted correlations.
- Correlations with the best battery.
- Number of misfits for each battery for the senior secondary group.

For ease of comparison we transformed all the results into comparable correlations by methods already explained.

TABLE XLIII

Collected correlations with success for the various tests and batteries

Battery	Duration in minutes	Senior secondary schools		Junior secondary schools		Misfits	Correlation with best battery	Average of (2), (4) and (6)	z
		Equal weighted	Best weighted	Equal weghted	Best weighted				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$IQ+Q+T_s$	..	·804	·808	·806	·808	·802	..	·804	1·11
$IQ+Q+S+T_s$	..	·800	·808	·799	·809	·802	·799	·800	1·10
$Q+S+T_s$	..	·790	·794	·784	·787	·793	·787	·789	1·07
$IQ+Q$	250	·786	·791	·774	·777	·792	·784	·784	1·06
$Q+T_s$	..	·783	·787	·780	·780	·780	·777	·781	1·05
$IQ+Q+S$	325	·783	·796	·773	·780	·779	·782	·778	1·04
$Q+S$	270	·774	·782	·753	·756	·777	·775	·768	1·02
Q	195	·770	..	·738	..	·760	·762	·756	·99
$IQ+S$	130	·736	·736	·741	·741	·736	·740	·738	·95
S	75	·698	..	·705	..	·701	·715	·701	·87
IQ	55	·691	..	·689	..	·690	·690	·690	·85
Fictitious batteries									
$IQ+T_s$	..	·779	·779	·795	·797	·787	·794	·787	1·06
$IQ+S+T_s$	..	·782	·785	·793	·802	·782	·787	·786	1·06
$S+T_s$	..	·764	·765	·775	·777	·770	·768	·770	1·02
T_s	..	·720	..	·741	..	·763	·724	·741	·95

We have labelled certain batteries 'fictitious' for the following reason. Teachers' estimates should not be used unless they are

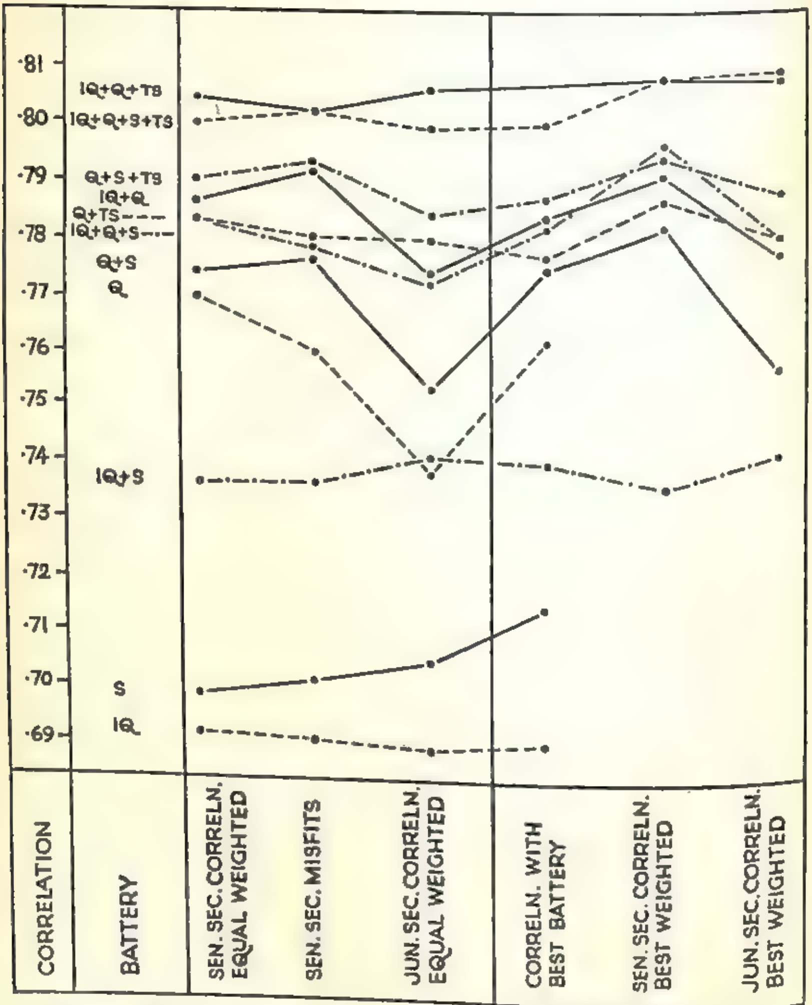

FIG. 42.—Graphical presentation of complete data as to order of merit of the various tests and batteries.

properly scaled on a uniform examination. They therefore imply that the results of a uniform examination are available, and would be used. We need, in fact, retain no battery which contains Ts unless it also contains Q.

In view of the results given in Chapter X it is unlikely that best weighting will be used in practice; and correlation with the best battery cannot be regarded as an authoritative test. We have therefore treated these merely as supplementary evidence, and used the averages of columns (2), (4) and (6) to give the final order of predictive value in column (8).

The final judgment is best made from the graphs in fig. 42, in which the determining measures are shown at the left and the supplementary ones at the right.

Conclusion

Fig. 42, the product of an immense amount of labour, is worthy of careful study. On all the determining tests IQ + Q + Ts comes out as the best battery, though its superiority over IQ + Q + S + Ts is slight. The difference may, indeed, be due to the fact that, in the complete battery, scholastic attainment is over-weighted in relation to IQ, an idea which receives support from the best weighted correlations.

After a distinct gap there is a set of four batteries between which the intervals are small. Yet the order is not in doubt. On all the determining criteria Q + S + Ts is the best of the four. IQ + Q, while somewhat lower in the correlation for junior secondary schools, is decidedly better than Q + Ts, which in turn is better than IQ + Q + S.

Thereafter the order is perfectly clear.

Turning now to the supplementary data on the right-hand side we find nothing to cast doubt upon our final order, and much to confirm it. The only discrepancy is the rise in standing of IQ + Q + S, due, in all probability, to the fact that scholastic attainment is over-weighted.

The order and spacing in fig. 42 conform closely to those in column (8) of Table XLIII, and we may take the latter as the basis for the final graphical presentation of the order in fig. 43.

As far as the order of merit of the single measures is concerned, our finding is that a carefully set and corrected examination is the best basis for the prediction of a pupil's success in the secondary school. Thereafter the order is, teachers' scaled marks, scholastic tests and intelligence quotient.

We have now attained one of our main objects, which was to

determine the order of predictive value of the various combinations of estimates, examinations and tests *as actually used.* This

CORRELATION	BATTERY	CORRELATION	BATTERY
·804	IQ+Q+TS		
·800	IQ+Q+S+TS		
·789	Q+S+TS		
·784	IQ+Q	·787	IQ+TS
·781	Q+TS	·786	IQ+S+TS
·778	IQ+Q+S		
		·770	S+TS
·768	Q+S		
·756	Q		
		·741	TS
·738	IQ+S		
·701	S		
·690	IQ		

FIG. 43.—Final order of predictive value of the various tests and batteries.

is the result which is of import to educational administrators; but, from other points of view, it is of interest to note the varying durations of the measures which, so far as they can be estimated, are given in the first column of Table XLIII.

SIGNIFICANCE OF THE DIFFERENCES IN THE FINAL ORDER

We have now to consider how far the differences in fig. 43 can be regarded as real, and not due merely to the accidents of sampling. This is no easy problem, but we give the following simple rule for finding whether the difference between the final correlations for two batteries is statistically significant.

If the difference between the z values[1] *for any two batteries is greater than ·04, the difference between the corresponding correlations is statistically significant.*[2]

The fact that not all the gaps are statistically significant is not surprising, and is no reflection upon the precision of our results. The more we increase the number of steps within a restricted range the more difficult it becomes to establish the significance of any one; and it must be remembered that while statistical significance establishes a real difference, its absence does not prove that there is no difference.

A large proportion of the mediate differences are significant. If, for example, an Education Committee is considering whether to use IQ + Q or IQ + S, it will find from the Table that the difference of z values is ·11, which is nearly three times the minimum for statistical significance. It is therefore left in no doubt that IQ + Q is the better battery.

We believe that the graphs in fig. 42 will carry greater conviction to the general reader than these statistical considerations. His confidence in the soundness of our final order will be determined by the fact that we have three different determinations, two absolutely independent and one partly independent, and that there are only two small deviations from the absolute conformity of all three orders.

[1] Given in column (9) of Table XLIII.

[2] If we have estimates of r_{QF} for two samples of n_1 and n_2 from equally correlated populations we can combine them by the z transformation. The combined result will have the same accuracy as a single correlation obtained from $n_1 + n_2 - 3$ pairs.

In our investigation we combined correlations such as r_{QF_1}, r_{QF_2} and r_{QF_3}, but these were determined on the same pupils. We cannot therefore maintain that the result is equivalent to a single correlation based on $n_1 + n_2 + n_3 - 6$ pairs. In calculating standard errors we counted these as different estimates of the correlation for the *same* sample.

In this way we obtained the standard error of the correlations for the senior and junior secondary schools. The correlations corresponding to the misfits are not wholly independent of those for the senior secondary schools, for they are based on the same group of pupils. They were, however, obtained by a different method, and we counted them as having the same standard error as the correlations for the senior secondary group. Knowing the three standard errors we then calculated the standard errors of the correlations in column (7).

The use of z values has the great advantage that the standard error of z is not dependent on the value of the correlation for the total population.

Suitability of the Various Batteries for Consideration of Border-line Cases: Effect on Order of Predictive Value

We have seen that by the aid of certain additional principles in border-line cases we can reduce the number of misfits for the complete battery by 6. Now, if we use IQ only we cannot adopt any of these additional principles; with a pair like IQ + Q, only a limited number. This consideration must therefore be taken into account in fixing the final order.

For the senior secondary group, omitting all pupils with absences in any item, we found the reduction in the number of misfits for all the batteries when the additional principles were applied. In so doing we introduced certain special rules to meet different cases. For instance, with a battery like Q + Ts, Tes and Tas were never counted as flops, but an examination mark was omitted if it fell below the corresponding teacher's scaled estimate by more than ·7 sigma.

The reduction in the number of misfits by the use of additional principles in border-line cases is as follows:

Number of tests in battery	Average reduction in number of misfits
4	4·0
3	2·7
2	·3
1	0·0

The reduction is greater with the batteries which include the larger number of tests, and a study of the complete data shows that the only effect on the order of predictive value would be to raise IQ + Q + S a little in the scale. Apart from that, the result would be much the same as that obtained by stretching the line in fig. 43, keeping the lower end fixed.

The detailed figures also suggest that additional principles should be used only with batteries of more than two tests.

Recommendations

We recommend:

That all pupils at the Qualifying stage be tested, not merely the applicants for bursaries or for admission to senior secondary schools.

That the battery used in urban areas be I + Q + Ts.

That, in rural and other areas where satisfactory scaling of teachers' marks is impossible, the battery used be I + Q.

Possibility of Escape from External Examinations

Our problem was to find the batteries of existing measures which are most effective in selecting pupils for secondary education. As scientists we should have had no feelings in regard to the answer; but, as educationists, alive to the evils of external examinations, our satisfaction at the consistency with which our experiments revealed the best batteries was not unmixed with regret that these all contained the Qualifying examination. Having stated our conclusions in a dispassionate scientific way, we may now give rein to our educational prejudices and consider whether there is any way of escape.

Perhaps the most promising suggestion is that we should use an intelligence test plus teachers' estimates scaled on its results. We thus almost completely avoid the evils of external examinations, but this advantage would be purchased at a cost.

In Chapter V we have given our reasons for believing that scaling upon intelligence score is unsound, of which the most important are:

With this method of scaling, the marks for all pupils from a particular school might be in error by as much as 12 per cent.

Since we should scale English and Arithmetic separately they would both have to be scaled upon intelligence score, and we know that the correlation between school means in English and Arithmetic is too low to justify this procedure.

The correlation with success of the best battery is ·804, and from fig. 43 it is clear that the correlation for the one now suggested will be less than +·787. Keeping in mind all the facts already given on this point, we estimate that the number of misfits in our secondary group of 416 (56 for the best battery) would be increased by about 5.

That is the debit side of the account; but against it we must set the educational advantages to the whole 3,000 pupils in the Qualifying group resulting from the removal of the incubus of external examinations from the primary schools.

Science is not yet capable of putting the credit side into statistical terms, and the decision must be left to the wisdom of the philosopher. But the problem is well worthy of further investigation, particularly in the direction of devising means of reducing the debit. This might be effected by such methods as the following:

Improvement of teachers' standards of marking, for example, by encouraging them to make more use of standardised tests.

Replacement of the present intelligence tests by some kind of composite test, which, while including English and Arithmetic, would be free from the evils of external examination.

Increase in the flexibility of the secondary system in the first year in such a way that transferences could be easily made and mistakes in selection rectified.

CHAPTER XVII

PROBATIONARY PERIOD IN THE SECONDARY SCHOOL

THE following results have some relevance to the last suggestion made in the preceding chapter.

PREDICTIVE VALUE OF MARKS IN FIRST YEAR OF SENIOR SECONDARY COURSE

Our method is to compare the prediction of the pupil's later success made at the Qualifying stage with that which we could make during and at the end of the first secondary year. The correlations between marks in the first secondary year and later success for the senior secondary group are as follows:

Average marks in	Correlation with average class marks in second year
1st term of 1st year .	+·854
2nd term of 1st year .	+·886
1st year . . .	+·894

These data show the extent to which our predictions would improve throughout the first secondary year, and give us a statistical basis for considering the best length of the probationary period.

COMPARISON WITH PREDICTION MADE AT QUALIFYING STAGE

For this purpose we take the results for the best battery, IQ + Q + Ts, and we must use the second-year average marks, for which, as we have seen, the correlations were rather low. This ensures that the ranges are the same for the correlations which we are comparing. Making also the allowance for the reductions in number of misfits by the use of additional principles in border-line cases (translated into correlation terms by the methods of Chapter XIV), we find that the comparable correlation is

$$r_{(IQ+Q+Ts)\,{}_{s}F_{Av}} = +\cdot 829$$

The position is therefore that with a good battery at the Qualifying stage we can reach a correlation of ·829; at the end of the first term of the first secondary year the figure is only ·025 higher. The difference is not so decisive as might have been expected when we consider how heavily the dice are loaded in favour of the later prediction. The decision is reached six months later, by which time the pupil has reacted to the new environment of the secondary school; it is based on a much larger number of tests, including those in the *new* subjects of the secondary curriculum. These are legitimate advantages of the prediction made at the end of the first secondary term, but there is a possible illegitimate one which arises from the halo effect.

Maximum Correlation with Success to be expected at the Qualifying Stage

One very interesting deduction which we can make from the above results is that no improvement of testing at the Qualifying stage is ever likely to produce a battery giving a higher correlation than ·854; for we could hardly expect that such a battery could yield a better result than a full set of examinations in the actual subjects and in the environment of the secondary school.

We feel, indeed, that we can, without incurring a charge of undue complacency, hold that our correlation for IQ + Q + Ts must be very near to the high-water mark of prediction at the end of the primary school course.

Minimum Number of Misfits to be expected with any System of Selection at the Qualifying Stage

By the methods described in Chapter XIV we find that if the number of misfits for our best battery were 53, then:

If the prediction were made on the basis of the first-term marks in the secondary course, this would be decreased by 4.
If made at the end of the second term the decrease would be 8.
If made at the end of the first year the decrease would be 9.

We can also prepare the following Table showing what the complete results of selection would have been if made at the end of the first term in the senior secondary school. It is applicable to conditions, likely to obtain throughout Scotland, where the distribution of applicants is skew and the proportion of unfit

candidates is 25 to 30 per cent. We have therefore given the figures in percentage form:

	Successes	Failures	Total
Admit .	66·9	6·9	73·8
Reject .	3·9	22·3	26·2
Total .	70·8	29·2	100·0

These figures are very important, in that they represent the best result that could possibly be obtained by any system of selection at the Qualifying stage; and they point to the following practical conclusions:

It is most unlikely that, under the conditions stated above, any system of selection at the Qualifying stage will ever yield results where:

(1) The number of misfits will be less than 10·8 per cent. of the applicants for admission to the senior secondary courses.

(2) Less than 9·3 per cent. of the number admitted will fail to make good.[1]

(3) Less than 3·9 per cent. of the applicants will be erroneously refused admission.

Application of Foregoing Results to the Problem of the Probationary Period

Let us consider first the suggestion that the probationary period should be one year. We should then have 9 fewer misfits (44 instead of 53) when the final selection of course was made—a very considerable saving. But, as usual, there is a debit side to the account. If, in the probationary year, we do not introduce the new characteristic subjects of the senior secondary course, we have no real test of the child's reactions to these subjects, and the prognostic value of the probationary year would be considerably reduced. If we do introduce them, we have to set against the saving of 9 misfits the serious wastage resulting from teaching these new subjects to a whole group, many of whom will give them up at the end of one year.

An exact assessment of this wastage is beyond our scientific power, but we would emphasise one point and make one suggestion.

[1] Assuming, of course, that the correct pass-mark has been used.

Whatever procedure be adopted in regard to a probationary period, we cannot escape the necessity for making as careful a prognosis as possible at the Qualifying stage; for we must then decide which of our Qualifying group are to get the chance of taking it. To give the chance to all would be sheer educational folly. If a careful selection for the probationary year is made, we suggest that a compromise procedure should be adopted in regard to the content of the course, giving only one foreign language. Thereafter, all that we can do is to try to ensure that change of course, or even of type of school, is as easy as possible.

Probationary periods of less than a year have been tried. On this point our results indicate that the superiority of prediction after one term over prediction at the Qualifying stage is hardly sufficient to warrant the dislocation of organisation which such an arrangement would cause. A period of two terms, while giving a better basis of prediction, is unlikely to be considered.

CHAPTER XVIII

THE MISFITS

AFTER trying all possible methods, and using all the supplementary data at our disposal in regard to health, industry and personal qualities, we were still left with 51 utterly refractory misfits.[1] There were 32 pupils who were failures in the senior secondary course, and, no matter how we used our data, they were admits; and there were 19 pupils who were rejects at the Qualifying stage, but who turned out to be successes. The time has now come to examine these with a view to finding pointers to the next steps in the improvement of our selection procedure.

THE ADMIT-FAILS

EXTENT TO WHICH THESE WERE BORDER-LINE CASES

One would expect that the majority of these final obstinate misfits would be pupils close to one or both of the admit-reject and success-fail border-lines; but an examination of their Qualifying scores shows that this is not so:

Qualifying score (IQ + Q + S + Ts) (sigma)	Boys	Girls	Total	Number who were doubtful failures in secondary course
+1·40–1·49	..	1	1	1
+1·30–1·39	2	..	2	..
+1·20–1·29	4	3	7	4
+1·10–1·19	2	..	2	..
+1·00–1·09	4	1	5	2
+ ·90– ·99	3	3	6	3
+ ·80– ·89	4	2	6	..
+ ·70– ·79	2	1	3	..
Total	21	11	32	10

There is certainly no evidence of bunching at the admit-reject line; and if we regard as border-line cases those falling between +·6 sigma and +·8 sigma, only three of the pupils would be counted border-liners on the entrance test. The last column

[1] This was the number for whom we had absolutely complete data.

indicates the number of cases, 10 in all, where the head teacher gave his judgment as to success in the secondary course with hesitation: all the others were perfectly clear failures. We may therefore take it that only 13 out of the 32 admit-fails fall into the category of border-line cases.

ILLUSTRATIVE CASES

The girl whose profile is given in fig. 44 is a perfectly clear admit on all tests; the home conditions are quite good, and neither the

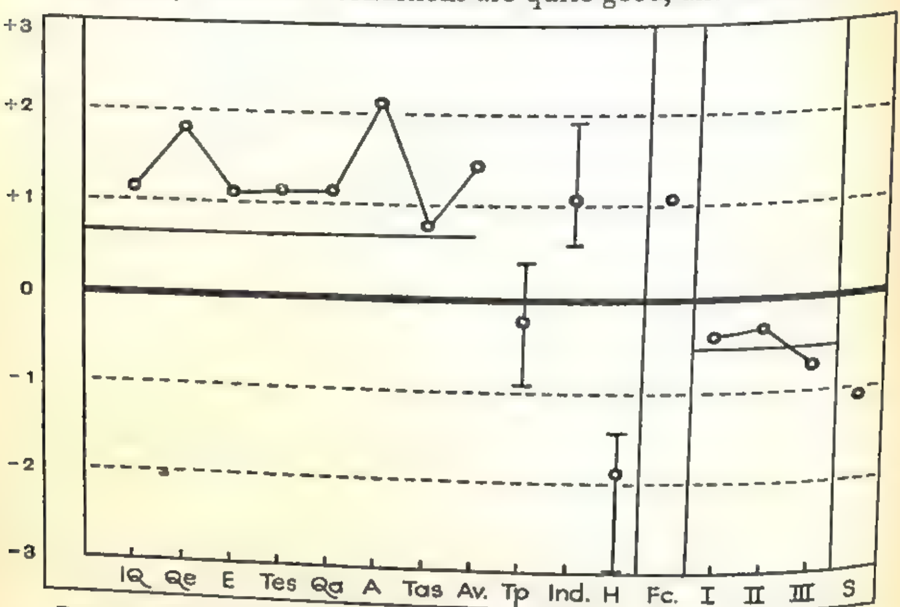

FIG. 44.—The admit-fail with the highest Qualifying score: Failure due to health.

primary nor the secondary head master has any fault to find with her attitude or interest. She herself was not particularly keen to take a senior secondary course; but she had no strong objection to meeting her parents' wishes.

We diagnosed this to be a case where the girl has been carrying on a losing struggle against indifferent health. At the Qualifying stage the Medical Officer's report recorded anæmia, defects of vision, and a general health grading of D. The secondary head master ascribes her failure largely to health reasons.

Yet, while we are fairly certain that taking account of health would, with this pupil, have saved us from error, we have seen that

as a general principle such a procedure would increase the total number of misfits. The only suggestion for improved selection arising from our consideration of this case is that where there is ground for serious doubt as to health the pupil concerned should be specially examined by the School Medical Officer.

We cannot see that we could have been blamed for admitting the boy whose profile is shown in fig. 45. There is not a single unfavourable indication in all his Qualifying data. His entry to the secondary school course seems to have been determined by the

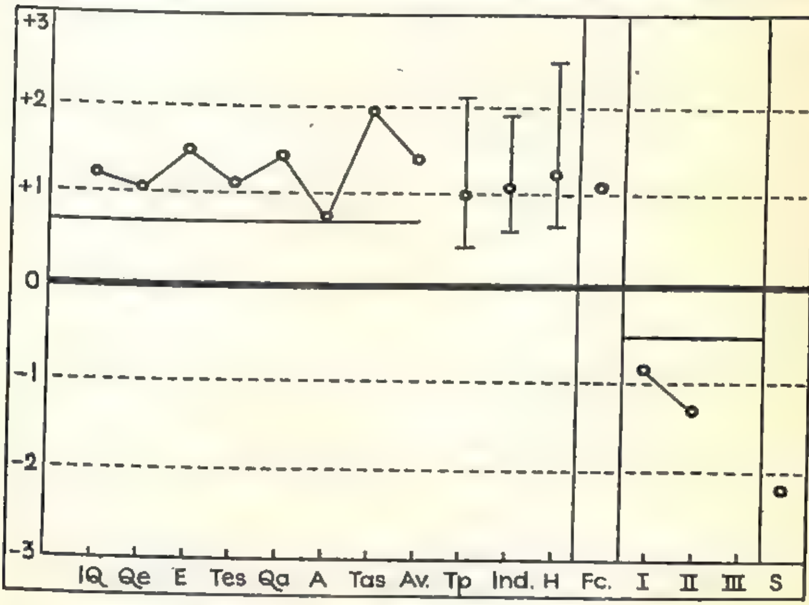

Fig. 45.—A failure due to home conditions.

fact that he was advised by his primary head teacher to sit the bursary examination. He was successful; and one gets the impression that it was a Scottish reluctance to give up the bursary rather than any keenness for secondary education that decided him to avail himself of it.

The view of the secondary head teacher was that the difficulty lay in the home. The boy got no encouragement from his parents, who seem to have soon regretted their decision to enrol him for the five-year course. As a consequence he had little interest in his studies and a strong desire to get away to work. One can well understand that this lack of home encouragement, while not

affecting the primary record, might have a more serious influence when the boy approached the leaving age.

In the puzzling case shown in fig. 46 something must have happened in the girl's history at the beginning of the second year in the secondary school. She is a perfectly clear admit on the test results, is above average in practical ability and in health, is graded A in industry; and the primary teacher thinks that she will be not merely a clear but a distinguished success in the senior secondary

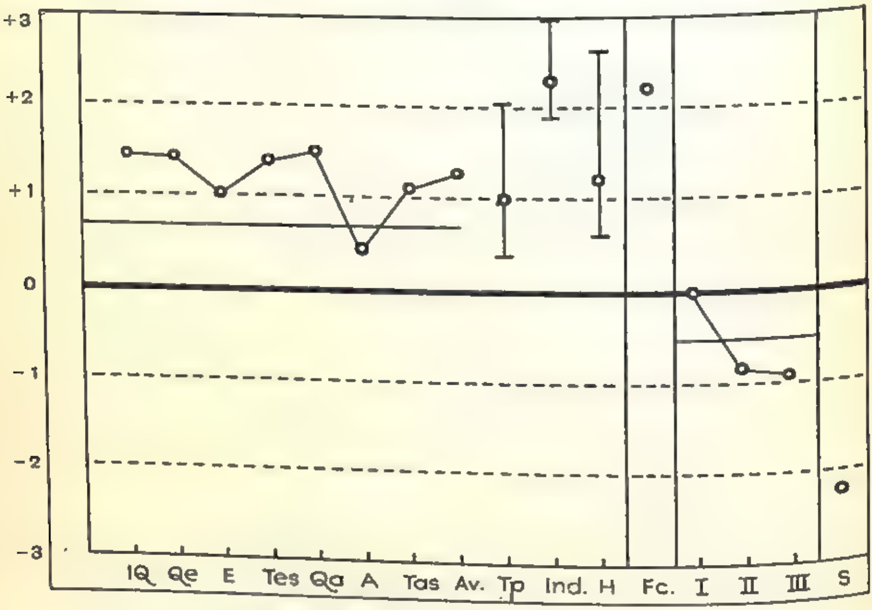

FIG. 46.—A failure due to emotional factors.

school. The primary report mentions perseverance and self-reliance as her marked personal qualities. The home conditions were favourable on the whole, but she was 'allowed to take life too easily.'

How then can we explain this remarkable drop to the level of a sure failure at the end of her first year? The only answer we can give is speculative. In discussing the pupil with the head master we were told that the drop was accompanied by signs of a definite change in personality, the details of which might give a clue to identification. The child was also handicapped by a certain deformity, perhaps not of much import at the primary stage, the

dawning sensitiveness to which might have serious effects on the temperament and character of an adolescent girl.

Perhaps the main pointer here is that such pupils should be referred to a Child Guidance Clinic. At any rate, no improvement in tests and examinations will save us from wrong placements in cases of this kind.

Fig. 47 gives the profile of a boy who is just above the admission mark but is a very clear failure in the secondary school. Two facts

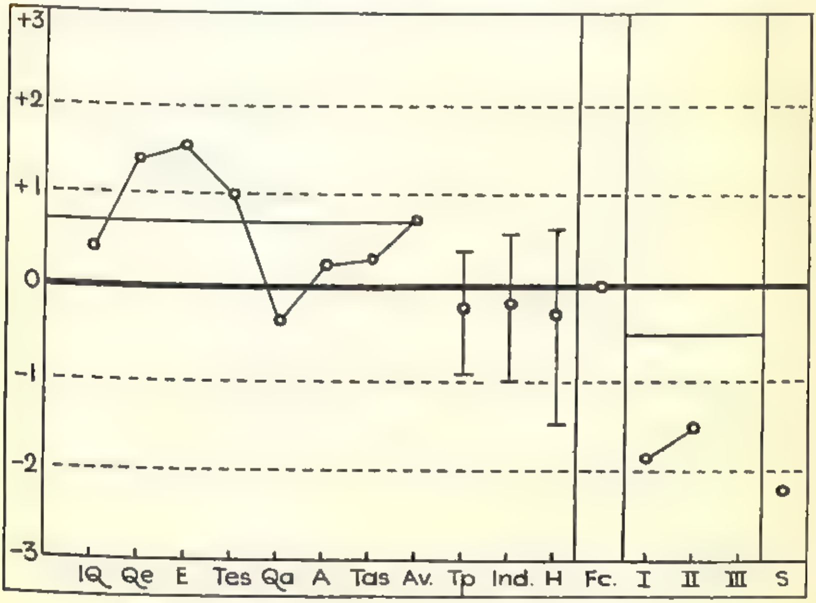

FIG. 47.—The admit-fail with the lowest Qualifying score.

seem to have contributed to his lack of success. In the first place, he was 'rather indifferent to school work' in the primary classes, his interests being mainly in games. He was not consulted in regard to his enrolment in a senior secondary course, went there with a grudge, objected to 'too much home work,' and left to become a salesman. The second cause was no doubt his weakness in Mathematics. Asked in Miss Rodgers's questionnaire[1] as to the subjects which he found difficult, he mentioned Mathematics, adding, 'It all depends on the teacher . . .' He must have had something else of interest to add here, for he had started another

[1] See p. 198.

clause, but unfortunately thought better of it. His profile shows that he is very good in English, but he just escapes the principle of the minimum of 0 sigma in Arithmetic. The clever boy can overcome such a weakness, but the mediocre or dull cannot; and continued failure and discouragement in one subject soon affect all the others.

The Reject-Successes

EXTENT TO WHICH THESE WERE BORDER-LINE CASES

The Qualifying scores of the reject-successes were as follows:

Qualifying score (IQ + Q + S + Ts) (sigma)	Boys	Girls	Total	Number who were doubtful successes in secondary course
+·90–·99	..	1	1	1
+·80–·89	..	..	..	..
+·70–·79	1	..	1	..
+·60–·69	2	1	3	..
+·50–·59	2	2	4	1
+·40–·49	3	1	4	3
+·30–·39	1	1	2	1
+·20–·29	2	1	3	2
+·10–·19	1	..	1	1
Total	12	7	19	9

The two pupils above +·70 sigma were rejects because they fell below the minimum of 0 sigma in IQ. The total number of border-line cases is 13.

ILLUSTRATIVE CASES

We now know that the pupil whose profile is shown in fig. 48 is a boy of intelligence and character, but we can see no way in which we could have admitted him on his Qualifying data. The primary teacher indicated self-reliance as a marked personal quality, and after reading his replies to Miss Rodgers's questionnaire we are inclined to agree. He did not sit the bursary examination, no doubt because financial assistance was not necessary. He concurred in his parents' view that he should take a senior secondary course, though one doubts whether the heads of the junior secondary schools would accept the implications of his wording. 'I entered a senior secondary course,' he says, 'because it is useful to have a good education in these days when one must be

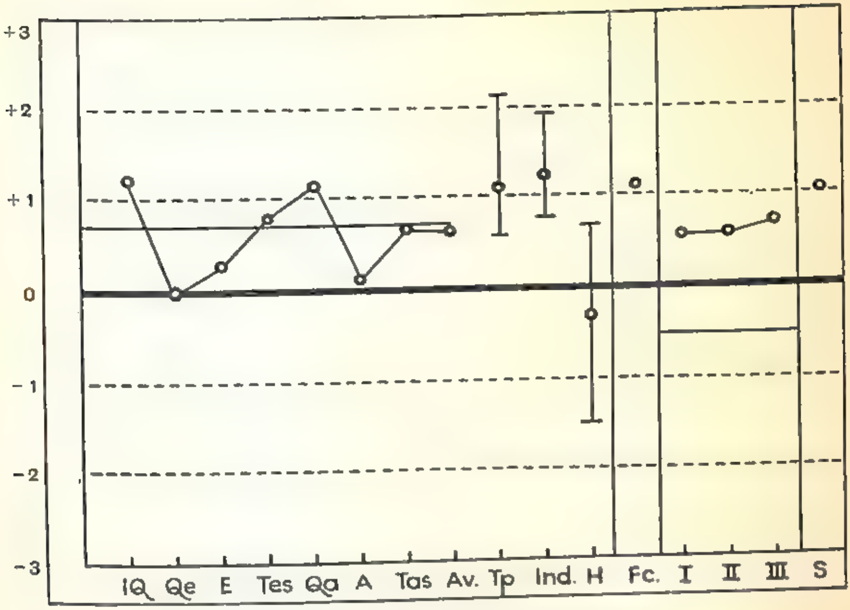

FIG. 48.—A border-line reject but a clear success.

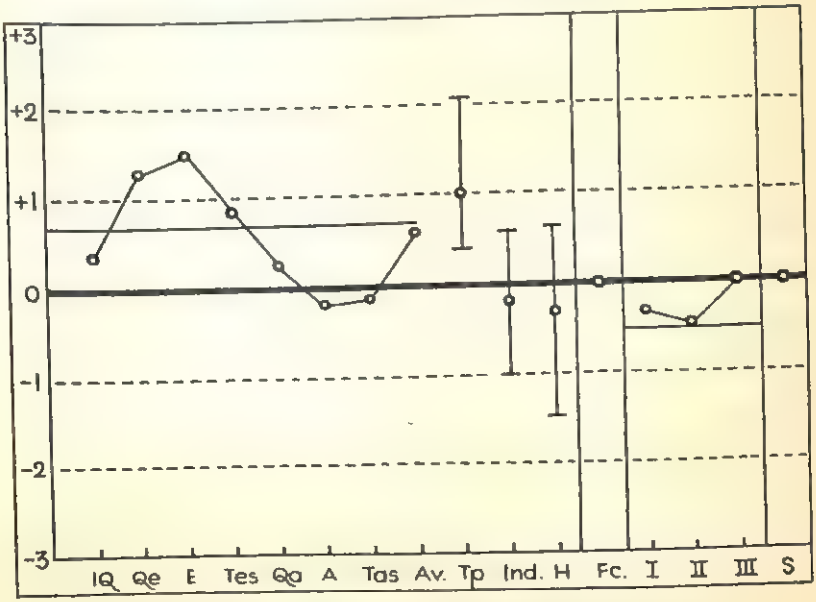

FIG. 49.—A border-line reject not much above the success line.

well-educated to find decent work.' Choosing the French-German course in preference to French-Latin because he thought 'a living language would be more useful in the present day,' he has found no reason to regret his choice. The main reason for his success in the opinion of the secondary head master lay in the home; the parents gave him every encouragement and were keenly interested in his progress. We are glad that our rejection of this boy was merely fictitious, and that he has had, and taken, his chance.

Fig. 49 illustrates the type of case which we should have expected to be much more common among those irreducible misfits. The boy is not far below the line for admission; indeed, if he had been just a little better in English he would have earned the lower pass-mark for special ability in one subject. He is also very little above the success line, and the head teacher gave his decision with some hesitation.

The boy's ambition is to become a doctor, but he is quite emphatic on the point that he is not in the right course. 'Two languages,' he says, 'are too much for me.' We hope he will win his fight which, but for good home conditions and encouragement, he would probably have lost.

Fig. 50 gives the profile of a bespectacled little boy from a good home who did not sit the bursary examination. In the primary school his interest in school subjects, with perhaps the exception of Music, was about average, and his attainments were well below the admission line. The choice of a senior secondary school was his own, though one gathers that he did take some account of his parents' wishes; and he is quite satisfied that it was a wise one.

At first we suspected that the remarkable improvement at the end of the first year might be due to late development, but the head teacher was of opinion that the change was on the side of personality rather than of intellect. Just about the beginning of the second year the boy seemed suddenly to acquire a sense of responsibility and then began to apply himself to his studies as he had never done before.

The type of case shown in fig. 51 constitutes a difficulty in follow-up work. The boy is a clear reject, and we have, from the prophetic point of view, no cause to complain of his first year's performance. On repeating the first year he jumped up to a clear success level, and maintained his position reasonably well in the

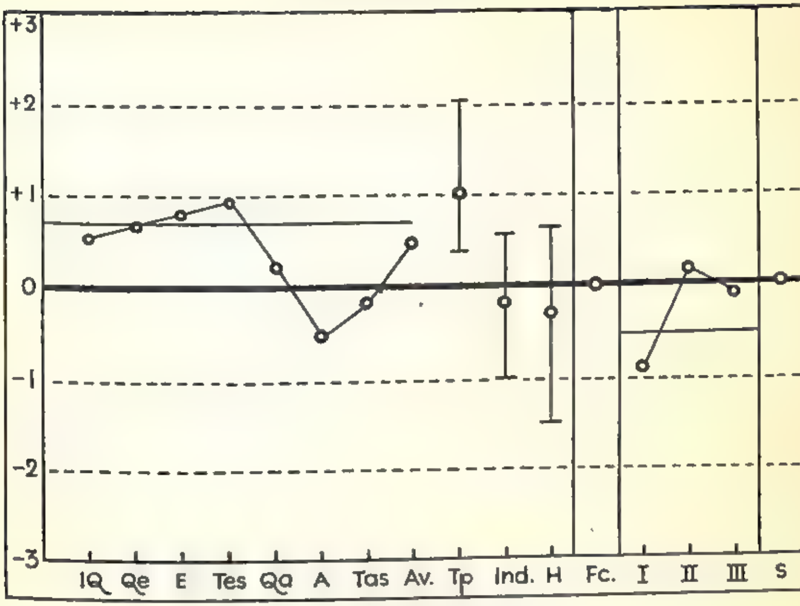

FIG. 50.—A possible 'late developer.'

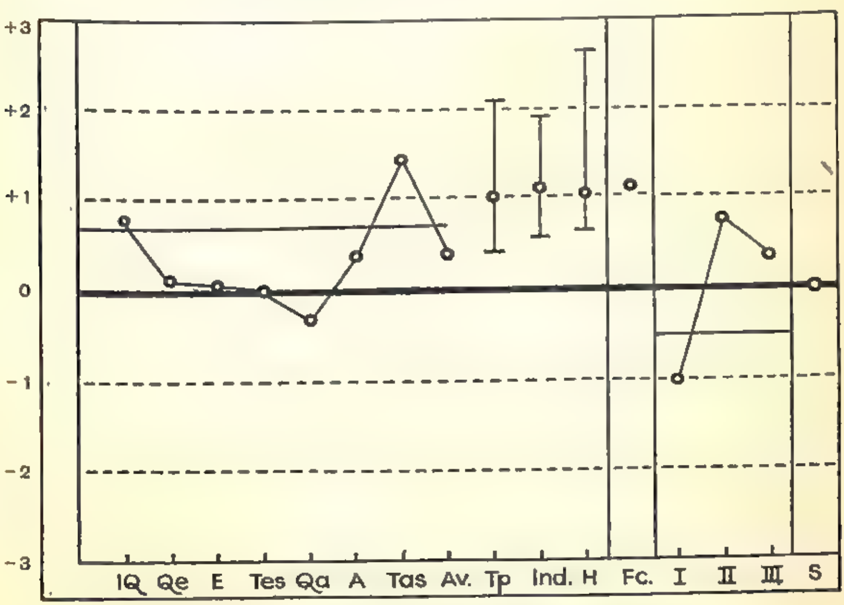

FIG. 51.—A pupil who repeated the first-year course.

following year. Taking account of all the facts, the head teacher was of opinion that he would stand the pace throughout the course and obtain the Leaving Certificate within the six years. He is a steady persevering boy who comes from a good and interested home.

Analysis of Causes of Discrepancy

PROPORTION OF CASES WHERE DISCREPANCY MAY BE DUE TO UNRELIABILITY OF ADMISSION AND SUCCESS MEASURES

Misfits may be due to lack of reliability of the measures or pass-marks for admission and success, or to extraneous factors like failure of health, deterioration of home conditions, etc. It is important to find out the relative proportions of these, for this will be our main pointer as to directions of further advance. If the former number is large we must strive to improve our tests; if the latter, we shall have fuller knowledge of the limits of any predictive procedure.

Possible cause of discrepancy	Admit-fails	Reject-successes	Total
Unreliability of measures			
Unreliability of entrance tests	10	4	14
Unreliability of success grading	3	9	12
	13	13	26
Extraneous factors (such as health)	13	5	18
No satisfactory explanation .	6	1	7
Total . .	32	19	51

The allocations were made with caution. To give an idea of the standards used, we may mention that the pupil of fig. 44 was classed as one where the success grading might be wrong; the pupil of fig. 47 as one where the entrance mark might be at fault. Further, unless we were absolutely satisfied that we had a convincing explanation on the ground of extraneous factors we relegated the case to the category, 'no satisfactory explanation.'

Of the 51 misfits 26 *might* be due to fallibility of the entrance or success measures: in 7 cases we have no satisfactory reason. These are almost certainly over-estimates; so we have full confidence in concluding that in *at least* one-third of the cases the discrepancy is due to extraneous factors and could not be avoided by any improvement of the tests and examinations.

EXTRANEOUS FACTORS CAUSING DISCREPANCIES

We selected in each of the 18 cases what we thought to be the main extraneous factor causing discrepancy and found the position to be as follows:

Cause of discrepancy	Admit-fails			Reject-successes			Total
	Boys	Girls	Total	Boys	Girls	Total	
Home conditions	3	1	4	1	2	3	7
Desire to take up work . .	2	1	3	..	..	..	3
Health . .	2	..	2	..	1	1	3
Unsuitable course	1	1	2	..	..	..	2
Personal qualities	1	..	1	1	..	1	2
Changes of school	1	..	1	..	..	..	1
Total .	10	3	13	2	3	5	18

In any one case, however, there are usually several contributory causes, and we present below the number of times that the various factors were mentioned in the reports. This gives a fairer measure of their relative importance:

Factor	Number of times mentioned		Total
	Admit-fails	Reject-successes	
Home conditions .	12	10	22
Unsuitable course .	10	..	10
Health . . .	6	2	8
Personal qualities .	6	1	7
Desire to take up work	4	..	4
Interest in a particular subject . . .	..	4	4
Changes of school .	2	..	2
Total . .	40	17	57

It will, of course, be understood that by health and personal qualities we refer to *changes* which took place after the Qualifying stage. The category, 'interest in a particular subject,' may be illustrated by a girl of indifferent Qualifying attainments who was talented in Art and desired to become an Art teacher. She realised that a successful secondary course was a necessary part of the road to her goal, and this provided an incentive which pulled up her work in all subjects.

Suggestions for Improvements in Methods of Selection

The smallness of our numbers, added to a consciousness of the deficiencies of our technique, precludes any definite conclusions. Yet our experience in working with these cases for nearly six years suggests the following directions of improvement in methods of selection.

Reports from the Primary Schools

We assume that the primary teachers will be encouraged to shoulder more and more of the responsibility for selection and guidance; and that, by a better technique than ours, they will be invited to give reports on home conditions, health, interests and personal qualities. In our view it would be both unnecessary and undesirable to ask them to make such reports on every applicant for admission to a senior secondary course. It is much better to have full and careful reports on a few selected cases than general reports on all. The knowledge that these reports would be carefully considered would increase their reliability.

Special Medical Examination in Doubtful Cases

If the primary teacher expressed doubts as to the child's physical fitness to face the strain of a senior secondary course, the pupil should be referred for special examination to the School Medical Officer.

Enlistment of the Resources of the Child Guidance Clinic

If the primary teacher was of opinion that the child's success might be prejudiced by the home conditions, these could be reported on by the Social Worker of the Clinic. Cases of difficulty in regard to personal qualities or cases of doubtful IQ could with advantage be referred to a Clinic Psychologist, who should also be consulted when a child suddenly changes in personality or in class standing during the secondary course.

Fuller Use of the Flexibility of the Regulations for Senior Secondary Courses

The number of misfits arising through unsuitability of the course will be reduced if the Education Committee uses a good battery of Qualifying tests and sees that the results are sent to the

secondary head teachers: but certain changes within the secondary school are also needed. In particular, there might be a reduction in the number of pupils who are encouraged to take two foreign languages, a greater provision of courses for adolescents whose cultural tastes are not met by the old 'normal general course,' and a greater effort to exploit the incentive that comes from special talent in subjects like Art and Music.

Extraneous Factors affect Boys more than Girls

One curious feature of our results is that in the misfits the boys always outnumber the girls, particularly in the admit-fail category. Here the ratio of boys to girls is 21 : 11 for the Senior Leaving Certificate, and 19 : 8 for the Junior Leaving Certificate. This seems to indicate that a girl tends to work up to her capacity better than a boy. It is possible, too, that a girl is not so liable to the discontent and unrest that come over so many boys in adolescence and lead to a desire to get away from school at all costs.

CHAPTER XIX

BURSARY EXAMINATIONS

THE object of bursaries for senior secondary courses is to ensure that no child who shows promise of profiting by such an education will be debarred therefrom by reason of the cost; and in the City where the Inquiry was conducted awards were, in 1935–1936, made on the basis of a special bursary examination. The numbers of candidates and of bursaries awarded were:

Group	Candidates	Bursaries awarded
December 1935	140	53
June 1936	275	130

IS A SPECIAL BURSARY EXAMINATION NECESSARY?

It has been a fairly common practice in Scotland to set a special examination on a higher standard for the award of bursaries. This was no doubt based partly on the belief that a stiffer examination would give a wider scatter of the good pupils, and ensure that, at the pass-mark, fewer candidates would have the same score. Clearly, if the examination were so easy that the clever pupils all scored full marks, this extreme ceiling effect would make the results useless for bursary purposes.

THE CEILING EFFECT

So far as we can see, the fear of a ceiling effect is no sound reason for a special bursary examination. In the first place, it is quite possible to have an examination for a complete Qualifying group which has neither a ceiling nor a floor effect. Our own examination showed neither; and reference to fig. 56[1] will show that this holds also for the distribution of scores on the complete battery. In the second place, the ceiling effect would be no disadvantage if the pass-mark fell some way below the point where the frequency was greatest.

[1] P. 197.

SPREAD OF CANDIDATES AT THE PASS-MARK

This consideration is certainly important; but it is rash to assume that a stiffer examination will necessarily effect an improvement; it may have the opposite effect.

While the fact that the bursary histogram in fig. 52 is lower on the scale is not conclusive proof that the examination was stiffer, a study of the papers confirms the belief. It will be noted that the frequencies at the verticals which cut off the actual number of

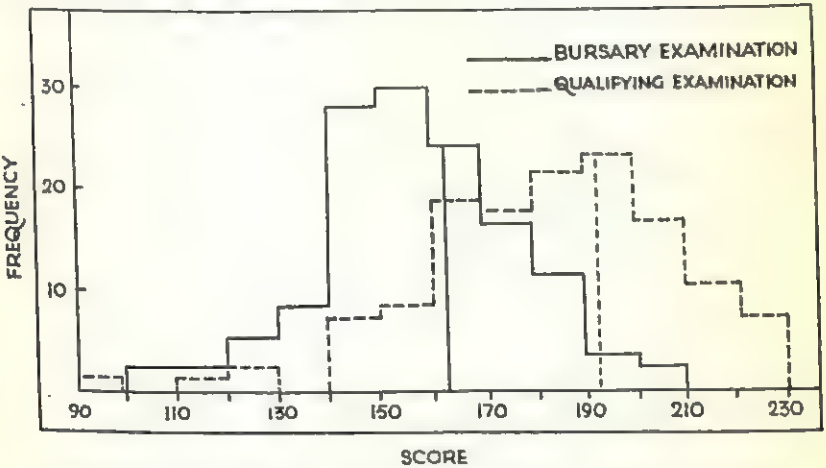

FIG. 52.—Score-scatter of December bursary candidates on the bursary examination and on the Qualifying examination.

successful candidates are very nearly the same in both examinations. The stiffer examination has therefore failed to secure a better spread at the critical point.

Fig. 52 shows how much of the effectiveness of the stiffer examination in spreading the candidates at the pass-mark depends upon the point at which the pass-mark is drawn. If it had been necessary to select less than 6 candidates, instead of 53, the stiffer examination would have been the better; but for almost any other number the easier examination would have been as good a basis of selection. Now, in Scotland it is never likely that only 6 candidates out of 140 would be entitled to bursaries; and we may therefore conclude that no statistical advantage is gained by a special examination of the standard used in December 1935. It is, of course, possible that a stiffer examination may give a better order of merit, though we think that this is unlikely.

Arguments against the Special Bursary Examination

Our view is that a special examination is neither necessary nor desirable; and that bursaries should be awarded on the basis of the Qualifying scores obtained by the application of a good battery to the whole Qualifying group. In arriving at this conclusion we had the following considerations in mind:

Qualifying scores *for the complete group* are required, not only for the award of bursaries, but also

(*a*) For guidance and selection for post-primary courses.

(*b*) As a basis for classifying entrants in the senior *and junior* secondary schools.

If teachers' estimates are to be used, as is very desirable, they must be scaled; and the scaling is possible only on the basis of an examination of the complete group.

Without Qualifying scores for the complete group there is no satisfactory means of fixing the pass-mark either for bursaries or for entrance to senior secondary courses.

A single examination is unjustifiable as a basis for the award of bursaries for secondary education.

Such an examination has no advantage over the Qualifying scores either in absence of ceiling or floor effects, or in spread of candidates at the pass-mark.

The institution of a special bursary examination offends the principle of 'parsimony of examinations,' which should have the same honour in the eyes of the educationist as 'parsimony of hypotheses' in the eyes of the scientist.

Correlations with Bursary Examination

Dr McIntosh calculated the correlations of the bursary examination marks with IQ and Qualifying score for the December and June groups, and found them to be as follows:

Correlation between bursary examination and	December (N=140)	June (N=275)
IQ	+·501	+·595
Q	+·591	+·544

The fact that the correlations with IQ are not so high as the corresponding figures for our own examination is partly due to the smaller standard deviation of the bursary group.

The most interesting point is that the correlations between the bursary examination and our own are as low as ·591 and ·544.

AN APPLICATION OF THE GENERAL PRINCIPLES OF CHAPTER XIV

With a correlation of ·591 between the two examinations the discrepancies would be very great, as the reader will by now realise. If, however, he has read Chapter XIV he should also be able to translate this indefinite statement into a fairly exact one; and by so doing he will have an opportunity of testing the usefulness of the methods therein described.

The following is a summary of the steps. We are selecting 53 candidates out of 140, so the number of discrepancies for zero correlation would be 66. From fig. 30 we see that the number for a correlation of ·6 would be 62 per cent. of this, namely, 41. That is, 20 candidates who get bursaries on our examination would be refused awards on the bursary examination; while 20 who were refused on our examination would receive awards on the bursary examination. The actual number was 19, showing that the method gives quite a good estimate.

If an Education Committee has ascertained the Number of Successes among the Candidates awarded Bursaries it can estimate the Number of Potential Successes among those who were refused

If the title of this section had not been so long we should have added 'and it would be most salutary for it to do so.' As a rule the candidates who are refused bursaries do not go into senior secondary courses. A large proportion of the mistakes made by the bursary scheme are therefore buried in the junior secondary schools; they do not obtrude themselves on the notice of the Committee, and this blind spot in its administrative procedure may give rise to an attitude of unwarranted complacency. It also gives the Committee little opportunity of knowing whether the pass-mark is fixed too high or too low.

The soundest method is, of course, a follow-up of the complete group of candidates; but this involves a considerable amount of labour, and a fairly reliable estimate can be obtained by the following short-cut.

After the bursars have been a year in the senior secondary schools the Director of Education would find out how many have been successes and how many have been failures. Using our December bursary group as an illustration, the numbers were 10 failures and 43 successes. Thereafter the procedure somewhat

resembles that used in the reconstruction of a prehistoric animal from a few bones.

All we know is the pass-mark AB (fig. 53), the numbers 10 and 43, and the total number of candidates, namely, 140. Our aim is to find an estimate of the number which should go into the quadrant BED. We must, of course, assume normality of the distributions, and we must have an estimate of the correlation between the bursary examination and success. The latter need not be very exact for

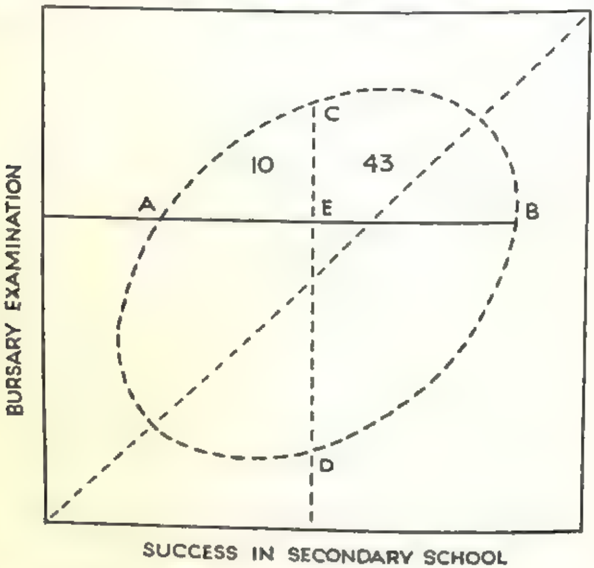

FIG. 53.—Estimate of number of successes among candidates refused bursaries.

this purpose, and the estimate could easily be made by any Director of Education who has been using scientific methods in his Area. In our own case the estimate would be $+\cdot 7$.

Using the notation of Chapter XIV, we know that

$$r = \cdot 7$$
$$k = +\cdot 3\sigma$$
$$\text{CEB} = 30\cdot 7 \text{ per cent.}$$

Now, knowing the volume CEB and the position of one of the planes which bounds it, namely, EB, it is possible by the method indicated in Chapter XIV to find the position of the other, namely, CE. We then find that $h = 0\sigma$, from which it follows that the number which should go into the quadrant BED is 27.

The conclusion is that 27 out of the 87 pupils who were refused bursaries would have made good in the senior secondary school. The number of correct awards was 43, and the number of mistakes 37. It will now be understood why we recommend this as a salutary type of calculation for Education Committees; but to allay possible suspicion of the mathematical character of the method, we proceed to give the actual numbers.

Analysis of Bursary Awards in the Light of later Success

From our follow-up data we know the number of successes and failures, both among those who were awarded and those who were refused bursaries.

two sample cases

We first give two sample cases in which we compare the predictions made on the basis of our complete battery and of the special bursary examination. Both are from the December group.

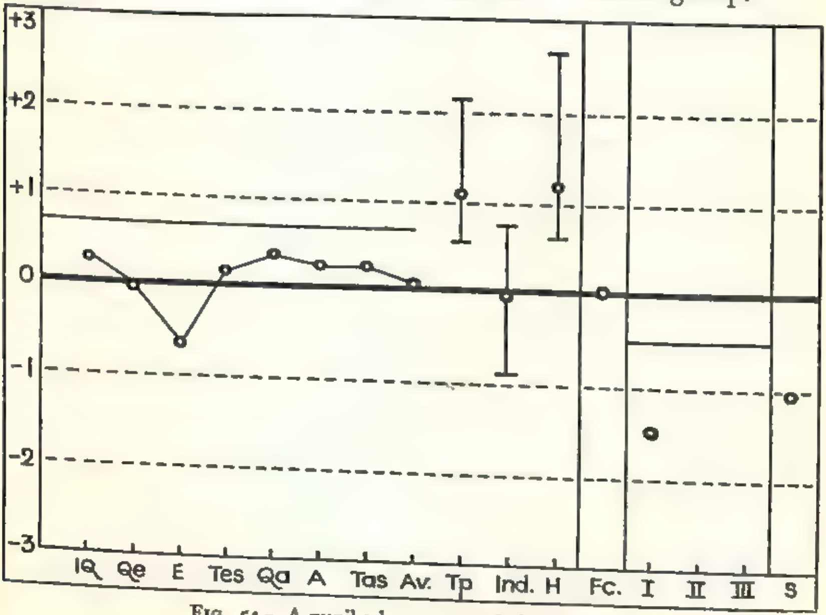

Fig. 54.—A pupil who was awarded a bursary.

The first, whose profile is shown in fig. 54, is that of a boy who was thirty-ninth on the bursary list and therefore received an award.

It will be noted that, on our Inquiry results, we should have had no hesitation in refusing a bursary to this pupil; and his performance in the senior secondary school shows that our decision would have been correct.

For comparison we give, in fig. 55, the profile of a pupil who was sixty-ninth on the bursary list and was therefore refused an award.

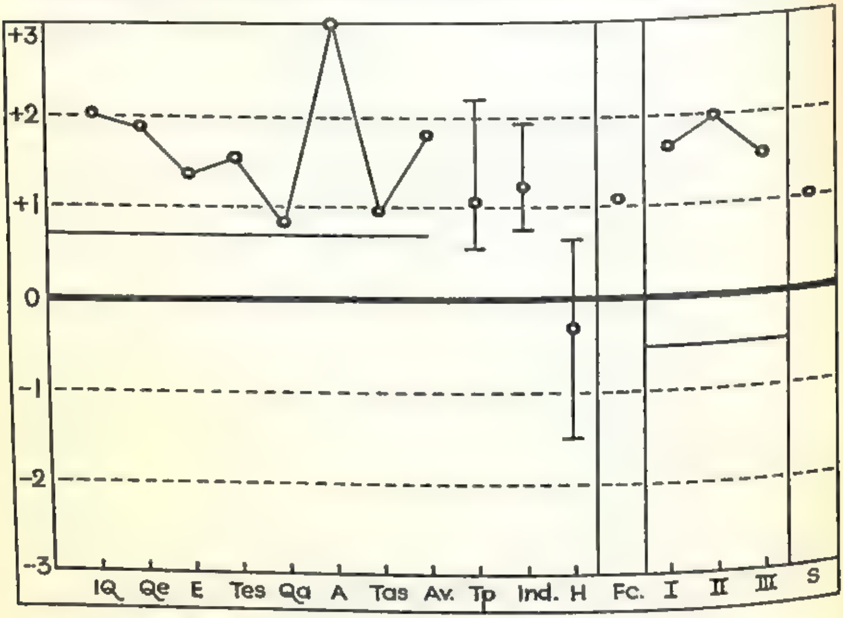

FIG. 55.—A pupil who was refused a bursary.

This boy, to whom we should willingly have granted a bursary, went on to a senior secondary school at his parents' expense.

COMPLETE RESULTS

The fitness of bursary candidates of both December and June groups for senior secondary education is shown below:

	Successes	Failures	Total
Awarded bursaries .	155	28	183
Refused bursaries .	69	163	232
Total .	224	191	415

The result is enlightening. Of the 183 candidates awarded bursaries 28 were unfit for senior secondary education; 13 of these

28 were, in fact, judged to be unfit even for a three-year junior secondary course. Even more disturbing is the fact that no less than 69 pupils who were fit to profit by a senior secondary education may have been debarred therefrom by reason of the cost. The total number of mistakes is 97 in all.

For the December group the actual number of refuse-successes is 26. The theoretical estimate in the last section (27) was therefore a close one.

We trust that the reader has by this time not merely a vague suspicion that the pass-mark in these bursary examinations was too high, but has proved that this *must* be so. Since the number of successes is greater than the number of failures, the number of award-fails should be greater than the number of refuse-successes at the correct pass-mark.

The Pass-Mark in Bursary Examinations

We shall assume that the battery I + Q + Ts is used and that the pass-mark should be determined solely on the basis of the fitness of the candidates to profit by a senior secondary education.

We have shown that, on this battery, a candidate for admission to the senior secondary schools has an even chance of success if his Qualifying score is +·70 sigma. It might therefore be expected that the same pass-mark should be used for the award of bursaries; but this does not follow. The position of the correct pass-mark depends on the nature of the group, its social class, home conditions, and so on. We found, for example, that the point at which the child in the whole Qualifying group has an even (statistical) [1] chance of success in the senior secondary course is not +·70 sigma but +1·34 sigma.

Now, for bursary candidates the extraneous conditions will be more favourable than in the complete Qualifying group, for in each case there is evidence of a desire for secondary education and the probability of home encouragement. Yet the home conditions and facilities for study will not be quite so favourable as in the group of candidates for admission to the senior secondary schools, many of whom come from homes which need no financial help. We should therefore expect that the pass-mark for the bursary candidates would lie somewhere between +·70 sigma and +1·34 sigma.

We determined the correct pass-mark (that is, the point at which

[1] See pp. 214–215.

half *of the bursary candidates* would be successes) by the methods already explained, and found it to be +·975 sigma. This means that the pass-mark used in the special bursary examinations was at least 6 per cent. too high. We explain in Chapter XXV how this pass-mark can be used in practice, but it should be remembered that it holds for the battery IQ + Q + Ts. With a poorer battery the pass-mark would be lower.

Results if Bursaries had been awarded on Basis of Best Battery with its Correct Pass-Mark

In awarding these bursaries we should have selected the battery IQ + Q + Ts; and, not having any financial responsibility, we should have used the correct pass-mark. Had we done so, the results would have been as follows:

	Successes	Failures	Total
Awarded bursaries .	192	41	233
Refused bursaries .	32	150	182
Total .	224	191	415

This should be carefully compared with the actual awards given on p. 194. Short-sighted ratepayers might be inclined to cavil at the expense of the 50 additional bursaries that we should have awarded; but, taking a wider social view, we find some compensation in the facts that we should have admitted 37 more pupils who would have been successful in a senior secondary course, and that we should have had 24 fewer mistakes.

We believe that these figures represent the best result that could be expected in the award of bursaries under present conditions.

CHAPTER XX

SURVEY OF THE QUALIFYING GROUP

FOR our survey we use as a basis the average standard scores on the complete battery IQ + Q + S + Ts, which we call the Qualifying scores.

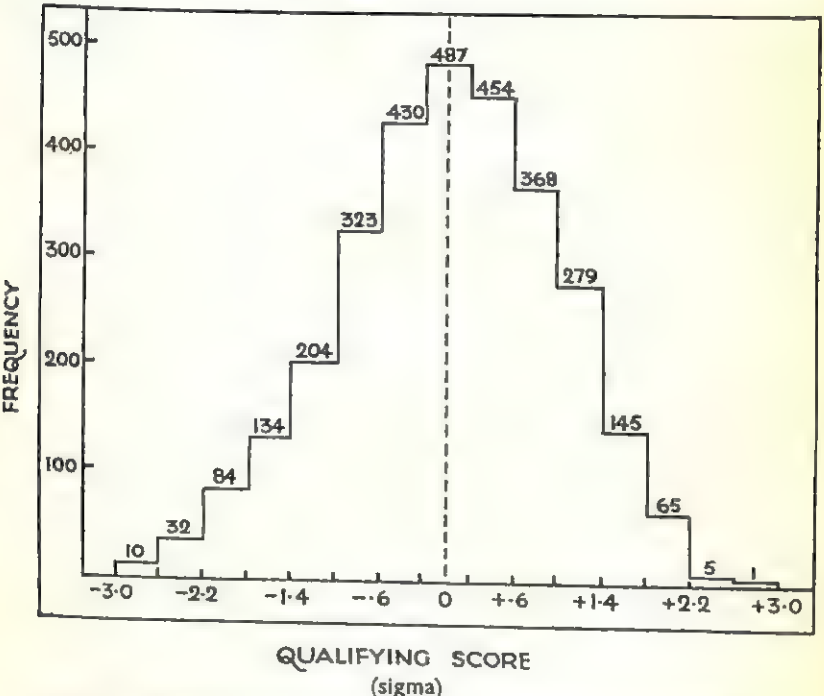

FIG. 56.—Distribution of Qualifying scores for complete Qualifying group.

COMPARISON OF BOYS AND GIRLS

COMPLETE QUALIFYING GROUP

There is no significant difference between boys and girls, either in mean score or in scatter. The difference of means is ·017 sigma [1] in favour of the girls, and the standard deviations are ·99 sigma and ·96 sigma for boys and girls respectively.

[1] This is about half the standard error of the difference.

SENIOR SECONDARY GROUP

The mean score for the boys is +1·018 sigma and that for the girls +1·208 sigma. The difference is more than three times its standard error, and is therefore statistically significant.

In an earlier investigation by Mr Robert Robertson it was found that the girls were better than the boys in intelligence and scholastic attainment separately, and that this result obtained not merely for the group as a whole, but in all but one of the individual schools. The school which was the exception was the smallest.

JUNIOR SECONDARY GROUP

In the junior secondary schools the boys are slightly better than the girls, the means being +·167 sigma and +·125 sigma. The difference is, however, not statistically significant.

If these tendencies should be confirmed in other years and Areas it would be of some interest to consider why the senior secondary courses should attract fewer of the clever boys than of the clever girls. Part of the answer may be that the roads to the vocations which attract the clever girl lie through the senior secondary school; but it may also be that the drags and pushes of the social environment, whose resultant is directed away from full secondary education, have more effect on boys than on girls.

Factors which Determine Choice of Secondary Course

(In collaboration with Miss Hilda C. Rodgers, M.A.)

The exacting nature of the main investigation did not permit us to make a full study of this interesting topic, but Miss Rodgers obtained some preliminary results from a questionnaire given to 391 senior secondary, and 420 junior secondary, pupils. Admittedly the problem cannot be solved in this way; the children do not know all the facts; we get little light on the motives which influenced the parents, or on the extent to which poverty debars pupils from senior secondary education.

CHOICE OF TYPE OF SCHOOL: SENIOR SECONDARY OR JUNIOR SECONDARY

One of Miss Rodgers's questions to the senior secondary pupils was, 'Why did you take a senior secondary rather than a junior secondary course?' and she gave a choice of three answers: '(*a*) Because I was advised to go to a senior secondary school by my

primary head master. (*b*) Because my parents wished me to go to a senior secondary school. (*c*) Because I myself desired to go to a senior secondary school.' She also left a space for 'Other reasons.' We thought that this would seldom be used, but it elicited a large variety of frank responses, given with a sublime disregard for the requirements of exact statistics. For many pupils the choice was determined simply by the fact that they had been in the primary division of a senior secondary school. As one casual young lady put it, 'I was in . . . and just carried on.' For others, the fact that a senior secondary school happened to be the one nearest to the child's home was the main consideration; quite a number of young educational oracles had the view that a senior secondary school gives a 'better education'; a fair number admitted candidly that they went to a senior secondary school simply because their friends had gone there; one or two, apparently with snobbish propensities, were most influenced by the fact that they would mix with pupils of a better social class.

Pupils were asked to indicate the main reason for the choice, and we give below what statistics we were able to extract:

Reason for choice of type of school	Senior secondary group	Junior secondary group
Advice from primary school	28	12
Pupil's desire	70	144
Parents' desire	196	95
Other reasons	78	121
Total	372	372

Even if we allow for their unreliability, the figures suggest that only to a limited extent is advice from the primary school a determining consideration in the child's choice of type of secondary school. Parents' ambitions and aspirations appear to be the main factor in the senior secondary group, and the pupils' own tastes in the junior secondary group. Confirmation of this was given in the answers to another of Miss Rodgers's questions: 'Were your wishes taken into account when your secondary course was decided upon?' To this question 94 per cent. of the junior secondary pupils gave an affirmative answer; in the senior secondary group only 73 per cent.

Of the 121 in the junior secondary group who gave 'other reasons,' 59 indicated that they wanted to leave school at the earliest possible date, and 20 gave as their reason that they had not been awarded bursaries. The others defied classification.

CHOICE OF COURSE WITHIN SENIOR OR JUNIOR SECONDARY SCHOOL

Miss Rodgers asked a similar question on this point which brought out the following replies:

Reason for choice of course	Senior secondary group	Junior secondary group
Advice from primary school .	15	12
Advice from secondary school .	69	7
Pupil's desire . . .	83	149
Parents' desire . . .	104	98
Other reasons . . .	89	73
Total . .	360	339

Again we see that the pupils' own desires have more influence in the junior secondary group and those of the parents in the senior secondary group. It is gratifying to find that advice from the secondary head master is an important factor in the latter.

That old traditions die hard is shown by the fact that no less than 3 pupils chose their course because of the 'value of Latin.'

TRANSFERENCES BETWEEN SENIOR AND JUNIOR SECONDARY SCHOOLS

Our results show that ease of transference from one type of school to another, when a mistaken choice has been made, is always likely to be an essential feature of an educational system. It is therefore interesting to find how far advantage is taken of such facilities for transfer as the present Scottish system offers.[1]

Miss Mitchell collected particulars of every change made by pupils in our Qualifying group and found the position to be as follows:

Not a single senior secondary pupil changed to a junior secondary school. Apparently when these children find that they are in the wrong school they simply leave at the minimum leaving age.

Only 12 pupils (6 boys and 6 girls) out of the 2,012 who went to junior secondary schools changed to a senior secondary school. All had spent a complete year in the first type, and all had to repeat the first year in the second type; so that for all these pupils there was a loss of a year.

[1] *Circular No. 44*, Scottish Education Department.

The point is of sufficient interest to warrant the presentation of Miss Mitchell's complete data:

Pupil No.	Qualifying score (sigma)	Course in Junior secondary school	Course in Senior secondary school	Length of course in senior secondary school	Success or failure in senior secondary school
1	+1·51	Commercial	French-German	Still on roll	Success
2	+1·28	Girls' technical	French	Still on roll	Success
3	+1·23	Commercial	French	1½ years	Success
4	+1·20	Commercial	French	Still on roll	Success
5	+1·19	Commercial	French	1 year	Success
6	+1·12	Commercial	French-Handwork	1 year	Success
7	+1·00	Commercial	French-German	Still on roll	Success
8	+ ·94	Commercial	French	Still on roll	Success
9	+ ·79	Commercial	French-Handwork	Still on roll	Success
10	+ ·65	Boys' technical	French	Still on roll	Failure
11	− ·05	Boys' technical	French	Still on roll	Failure
12	− ·02	Boys' technical	French	1 year	Failure

The Qualifying score is that on our complete battery, and it is comforting to note that with our pass-mark of +·70 our prediction would have been correct for every one of these pupils. We should certainly have discouraged the last two from making the change. Nine out of the 12 have made good in the senior secondary school, and 6 are likely to complete the course. In the light of these figures it would seem that the flexibility of our system is not so great as is sometimes thought.

CHANGES OF COURSE WITHIN SENIOR AND JUNIOR SECONDARY SCHOOLS

Miss Rodgers's questionnaires were given 1 year and 8 months after the pupils had started the secondary courses, and certain of her questions bring out the flexibility of the system within the two types of school.

She found that of the 391 pupils in her senior secondary group, 58, or 14·8 per cent., had changed from one course to another, the main change being from a two-language to a one-language type. No less than 76 per cent. of the 58 changes were of this kind; 9 per cent. changed from one to two languages; and the

remainder were mainly due to the dropping of technical or domestic subjects.

Another interesting point was that about one-fifth of the pupils were still dissatisfied with the course they were in. Most of these discontents were in the two-language courses and felt that they would have been happier and more successful if they had taken a technical or commercial course, usually the latter. Indeed, 15 pupils, who were satisfied with their choice, added a note to the effect that their literary courses would have been improved by the addition of some commercial work. This adding of outspoken critical notes was a common feature of the questionnaires; and we read the pupil's educational views with an interest not unmixed with admiration for their soundness.

These results confirm the conclusions recorded later as to the number of pupils who should be encouraged to take two-language courses. They also suggest that the slow growth in popularity of the newer types of secondary course is not the fault of the pupils. Two further points that may be of interest to those who plan the courses are that a large number of senior secondary pupils taking a commercial course expressed the view that commercial subjects should be started in the first year, and that many pupils chose a junior secondary course rather than a senior secondary commercial course because, in the former, they could begin the commercial subjects earlier.

Of the 420 pupils in the junior secondary group only 17, or 4 per cent., had changed their course. Practically all possible transfers were represented, the commonest being from the commercial to the technical. Seventeen pupils were still dissatisfied with their course.

Many of the facts elicited by Miss Rodgers's questionnaires have a bearing upon the problem of the omnibus school.

Comparison of the Four Groups at the Qualifying Stage

qualifying scores

The four groups were senior secondary, junior secondary, backward and retained, and the distributions of the average scores on our complete battery are shown in fig. 57.

These figures show the way in which the nation's resources of talent distribute themselves when almost free choice of school and course is left to parents and pupils. They are puzzling as well as illuminating.

The broken vertical is drawn at +·70 sigma, the point above which the candidates for a senior secondary course have more than a 50 per cent. chance of success. We may therefore assume that all pupils above this line have the ability to profit by a senior secondary course. A glance at the histograms then shows that, while the 6 best pupils have all gone to the senior secondary school,

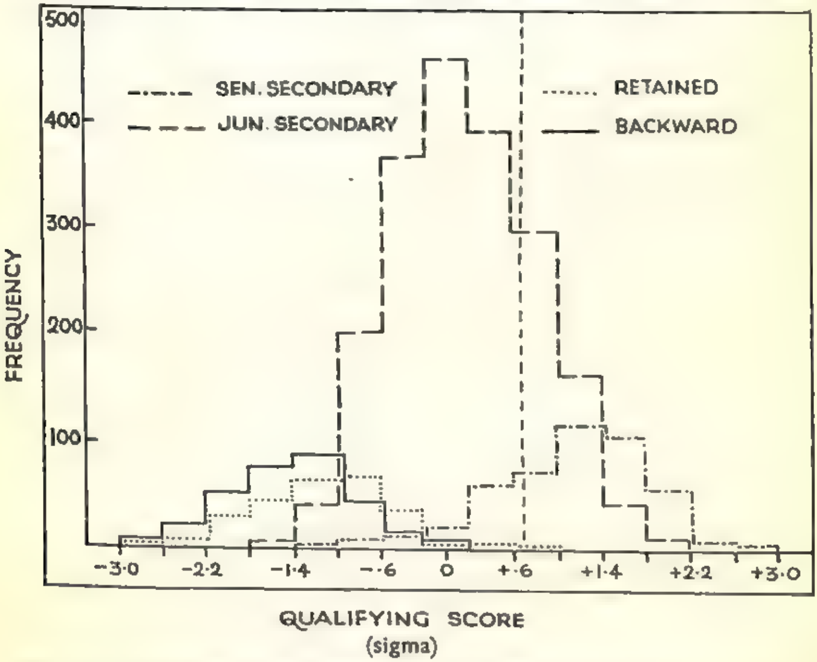

FIG. 57.—Distribution of Qualifying scores for the senior secondary, junior secondary, backward and retained groups.

a very large proportion of those who had the necessary ability have not taken advantage of this educational opportunity.

The next striking point is the tail of the senior secondary group and the distance it stretches below the success border-line, penetrating to near the bottom of the junior secondary group and to about the mean of the backward group.

The lower part of the distribution of the backward group is in conformity with expectation, but those in the top intervals would seem to have deserved a better fate. Why, one wonders, should 6 pupils with attainments up to the average of the whole Qualifying group be sent to these special classes? For only one did we find an explanation: the girl simply refused to go to a junior

secondary school. Of the other 5, one was promoted to a junior secondary course, the others were retained in the backward classes, where the head teachers felt that their educational needs were satisfactorily met.

The overlapping of the retained and secondary groups is almost incomprehensible. Three pupils who fall within the top 23 per cent. of the whole group are retained, while 3 who fall into the bottom 16 per cent. are sent to senior secondary courses.

The means and standard deviations of the Qualifying scores of the four groups were as follows:

Group	N	M	σ
Senior secondary .	452	+1·102 sigma	·67 sigma
Junior secondary .	1,975	+ ·146	·65
Retained . .	278	−1·167	·65
Backward . .	316	−1·429	·58

INTELLIGENCE

In an early preliminary study made before the data for the retained group were available, Mr Robert Robertson found the following results for intelligence:

Group	Mean IQ
Senior secondary . .	115·7
Junior secondary . .	102·0
Backward . . .	84·0

HEALTH, INDUSTRY AND INTERESTS

Mr Robert Robertson found the percentages of the groups who were awarded the gradings A, B, C, D and E in each of these factors. To reduce the results to a simpler and clearer form we calculated average scores by giving five points for an A, four for a B, and so on:

Group	Health	Industry	Interests		
			Humanistic	Realistic	Practical
Senior secondary	3·32	3·66	3·63	3·69	3·48
Junior secondary	3·22	3·22	3·20	3·24	3·24
Backward. .	3·13	2·30	2·15	2·16	2·47

These results have not the reliability which would warrant an attempt to calculate the statistical significance of the differences

shown. Yet the figures are not without interest, whether we compare the rows or the columns. In interpreting them it is well to keep in mind that the score for a normal distribution would be 3·00.

In health the senior secondary pupils are the highest, and the standard falls progressively from the junior secondary to the backward groups, a result which bears out the view that intellectual and physical fitness are associated. This is probably the most reliable of the results, for it was based on careful comparable gradings by the School Medical Officer. For industry, where the fall from group to group is more striking, we have to realise that it must be difficult for the teacher to grade without being influenced by the pupil's school attainment. Also, the disentangling of cause and effect would be possible only with a much more careful technique than we were able to adopt.

We have already given reasons for distrust of the reliability of the interest gradings, but from the first row we see that the predominant preference of the senior secondary group is not for the humanistic but for the realistic. This may bring some encouragement to those who believe that the curricula of these schools are too literary and linguistic, and that realistic and practical subjects should have a more prominent place in some of them. These senior secondary pupils are, indeed, more interested in practical subjects than the pupils in the junior secondary group, whose tastes are more evenly distributed. On the whole, however, the interests of the latter are rather realistic and practical than humanistic.

The interests of the backward group presumably lie elsewhere than in the world of the school, for they are well below 3·00 in all types of subject. Yet they make their highest score in the practical group, which supports the belief that the curriculum for pupils of this type should consist largely of manual activities.

Comparison of the Sections taking different Courses in the Senior Secondary Schools

The main interest of this comparison is that it allows us to judge how far the newer senior secondary courses had, by 1935, begun to attract the abler pupils. When the Latinless section of the French *lycée* was abolished in 1923, a *professeur* protested that this section had formed a convenient depository into which, with a

light heart, were dumped the pupils with insufficient ability to do well in classics: and he asked petulantly, 'What is now to be done with these undesirables?' To find whether the modern and technical courses perform the same service in Scottish schools we prepared the following comparison:

Section	N	Mean Qualifying score	Mean IQ
Latin-French . .	151	+1·367 sigma	120·0
French-German .	110	+1·315	119·0
French-Technical .	74	+1·006	114·7
French-Literary .	100	+ ·619	108·0

In the French-Technical category we included French-Domestic and French-Commercial, and it is clear that such courses are beginning to attract a fair proportion of the clever pupils. The dumping-ground for weak entrants would appear to be the French-Literary course.

Comparison of Sections taking Boys' Technical, Girls' Technical and Commercial Courses in the Junior Secondary Schools

The mean IQs and Qualifying scores for the junior secondary courses were as follows:

Course	Qualifying score		IQ	
	M	σ	M	σ
Commercial .	+·681 sigma	·59 sigma	110·2	10·3
Boys' technical	+·078	·60	101·2	10·3
Girls' technical	−·051	·60	98·7	9·5

The commercial course is apparently the one which attracts the best of the junior secondary pupils. It is usually divided into two sections, those who take French and those who do not; and Mr Robert Robertson finds that the average IQs of these sections were 114·4 and 106·1. The commercial group in the junior secondary schools is therefore somewhat better in quality than the French-Literary group in the senior secondary schools; and that part of it which takes French is of the same quality as the French-Technical group.

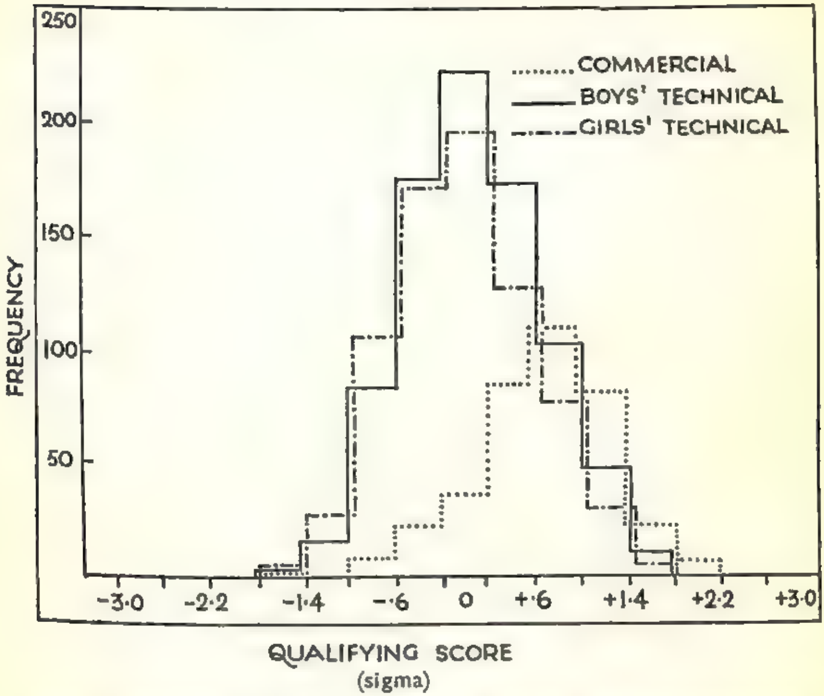

FIG. 58.—Distribution of Qualifying scores for courses in junior secondary schools.

SYSTEM OF RETENTION

The retention system is a departure from the 'clean cut,' which provides for the pupil who, if promoted, would have to go to a backward class, but who would be fitted for a normal secondary course if his attainments were increased by a further period in the Qualifying class. Our data provide some information as to its practical operation.

Below we give details of the duration of the period of retention before promotion for the retained pupils in our group:

Retained for	Number who were fit for Senior Leaving Certificate	Junior Leaving Certificate	Day School Certificate (Lower)	Shorter course	Fit only for backward class	Total
1 year 6 months	..	..	3	3	..	3
1 year	1	6	23	54	43	97
6 months	1	3	16	80	77	157
No post-primary career	..	..	..	..	21	21
Total	2	9	42	137	141	278

The following were the courses taken by pupils after retention:

Course taken	Number who were fit for Senior Leaving Certificate	Junior Leaving Certificate	Day School Certificate (Lower)	Shorter course	Fit only for backward class	Total
Senior secondary .	1	3	6	6	..	6
Junior secondary .	1	6	36	108	1	109
Backward . .	..	..	..	23	119	142
No post-primary career . .	..	..	..	..	21	21
Total .	2	9	42	137	141	278

The second set of data shows that the system serves a useful purpose. If we assume that all the retained group would have gone to backward classes in a clean-cut system, we see that it has saved about half of them from that fate. While many of these would no doubt have been transferred later from the backward group, the advantage of keeping them in a normal Qualifying group is considerable.

IMPROVEMENT AFTER SIX MONTHS' RETENTION

We present fig. 59 partly to give an idea of the improvement in standing resulting from six months' retention, and partly as an example of a case where we could not see why retention was thought necessary. This girl, who was first presented at 11 years, falls above our pass-mark for admission to a senior secondary course. She is described as persevering and self-reliant, and the report on the home is 'considerable parental interest.'

After six months' retention she improved her Qualifying score from +·82 sigma to +1·20 sigma, and then proceeded to the commercial course at a junior secondary school. Thereafter, she transferred to a senior secondary school where she had to take the first-year course, thus losing one and a half years in all.

RECOMMENDATIONS WITH REGARD TO RETENTION

A detailed study of the data suggests the following tentative conclusions regarding the system of retention:

The system of retaining pupils for six months is valuable in certain cases.

Retention for one year should be resorted to only in exceptional circumstances, and retention for longer periods should be abandoned.

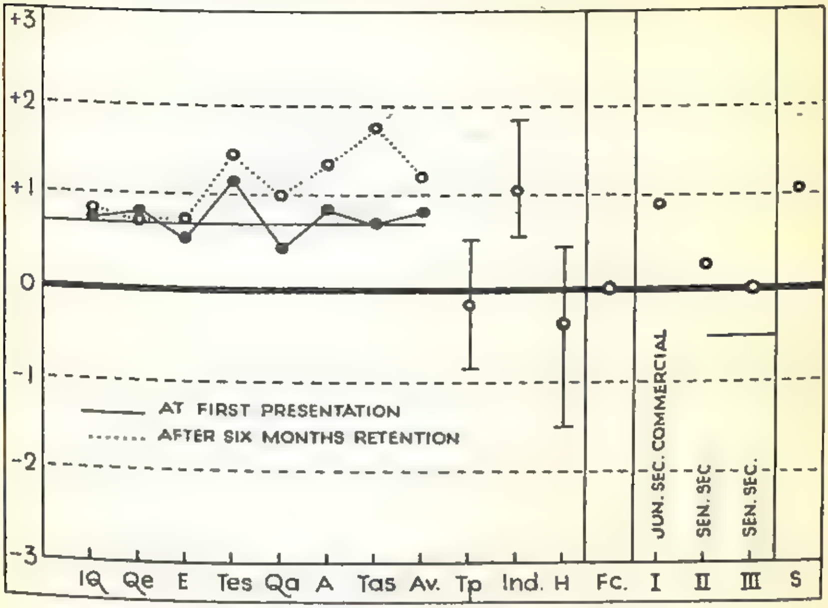

FIG. 59.—Case where retention was probably unnecessary.

Another striking point is that the number who, after retention, are found to be fit only for backward classes is rather high. If these pupils could have been detected at the first presentation they could have been sent to the backward classes at once, thus relieving the Qualifying teachers of some dead-weight, and giving a sounder educational treatment to the pupils themselves. It is certainly not to their educational advantage to keep them at a hurdle which is too high for them, and run the risk of further sapping their self-confidence. This is an aspect of the selection problem which would repay further attention by Education Committees. Like many others, it raises the question of the desirability of a preliminary sifting of the material at an earlier stage in the primary course and of a development of special classes within the primary school.

If, in consultation with their head teachers, Committees will fix their policy in regard to backward classes, the type of course to be offered, the type of pupil who should be sent to them, and so on, it would be possible to fix a border-line as for senior secondary courses. At the moment the attitudes of the head teachers to these classes vary so much that the problem is insoluble.

The Backward Group

The distribution of Qualifying scores for this group, shown in fig. 57, would lead us to suspect that a number of pupils had been relegated to the category in error. This is confirmed by the judgments of the teachers who taught them in the backward classes:

Judgment of teachers as to degree of fitness of pupils sent to backward classes	N
Fit only for backward class	283
Fit for junior secondary course after retention . .	26
Fit for junior secondary course at time of transference to backward class	7
Total	316

The number of errors in placement, while considerable, was therefore not unduly large.

CHAPTER XXI

PERCENTAGE OF PUPILS FITTED FOR (*a*) A SENIOR SECONDARY EDUCATION, (*b*) A JUNIOR SECONDARY EDUCATION OF JUNIOR LEAVING CERTIFICATE LEVEL

This old problem, whose obituary notices have so often appeared in the writings of educationists, has an administrative vitality which has preserved it amidst the constantly changing ideas of secondary education. But present nomenclature makes it difficult even to state precisely what the problem is; and any answer to it is destined soon to go out of date. All we can do is to give estimates for secondary courses *of the types that are offered now and with the standards that now obtain.*

SENIOR SECONDARY EDUCATION

Distinction between Potential and Actual Successes

From our data we can estimate the percentage of the Qualifying group who would actually have been successful in a senior secondary course. The difference between this and the percentage who actually took and succeeded in such a course is a measure of what may be regarded as waste of national talent.

In assessing the waste, from the point of view of the individual pupil, we must bear in mind that the difference is mainly in length of course, and that the needs of many clever pupils may be best met in the junior secondary school. But from the statistical standpoint the measure is too lenient.

A TYPICAL CASE

In fig. 60 we give the profile of a boy with a Qualifying score of +1·80 sigma who went to a junior secondary course and was judged by his head master as one who would have been a certain failure if he had gone to a senior secondary school.

The head master's view that this boy would have been a failure in a senior secondary course was based upon knowledge of his unfavourable home conditions, where there was a constant struggle

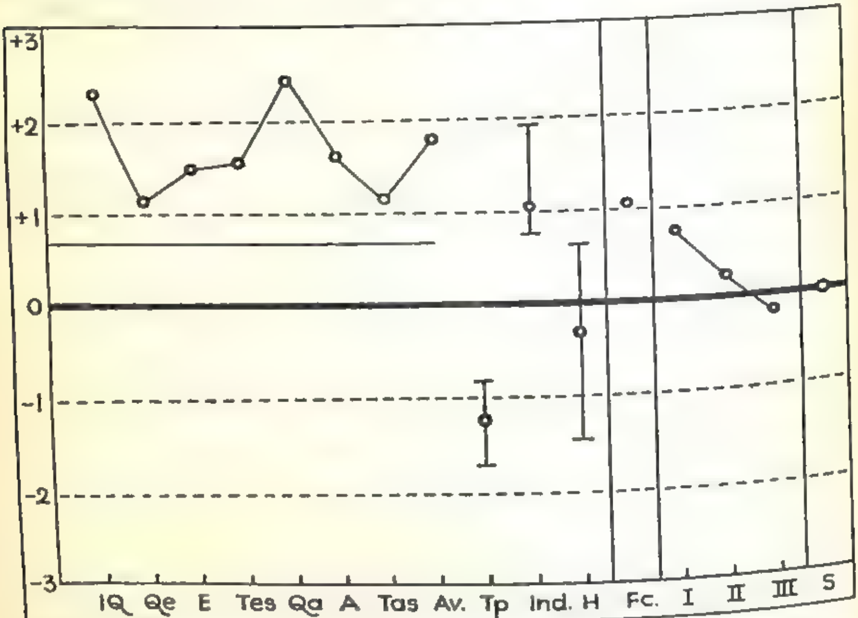

Fig. 60.—A pupil who had the ability and attainment for success in a senior secondary course but who would have been a failure.

with poverty and where he received no encouragement in his studies. In interpreting the profile it has to be remembered that the follow-up marks are relative to the mean of his *junior* secondary course.

ANALYSIS OF 78 CASES WITH QUALIFYING SCORES OF +1·1 SIGMA AND OVER

In discussing individual pupils with head masters of the junior secondary schools we frequently came upon instances of the above type. The child had the necessary ability and attainment but would have succumbed to the resultant drag of his environmental conditions. Sometimes the source of the trouble lay in the poverty of the home, and the parental eagerness for the child to make his contribution to its meagre resources. In such conditions the child had no home encouragement in his further education;

often a positive antagonism, the offspring of sheer necessity. Frequently we were told that the child was a member of a large family living in a house affording no facilities for the home study demanded of the senior secondary pupil. Even when the home conditions were moderate or good, other strong forces of the environment seemed to be arrayed against the pupil; his friends were all leaving school at 14 to become van boys or message boys, with their evenings free and a little money in their pockets, and he wished to leave with them. The attraction of the senior secondary school was far too weak to overcome these opposing pulls.

In the junior secondary courses there were 78 pupils with Qualifying scores of +1·1 sigma or over who would have been failures in the senior secondary school if they had gone there. Six out of 7 pupils in the senior secondary group with this Qualifying score are successes; so we may take it that these 78 pupils had the necessary ability. Their cases were made the subject of independent study, and the following reasons why they would not have been successful in the senior secondary school were found:

Reason for failure	N
Unfavourable home conditions . . .	18
Lack of interest : desire to take up work .	16
Personal qualities : laziness, delinquency, etc.	8
Ill-health.	4
No suitable course in senior secondary school	3
No known reason, apart from the fact that the pupil's performance in the junior secondary school was unsatisfactory . . .	29
Total . . .	78

The categories overlap; and when more than one cause was operative we placed the pupil in accordance with what the head teacher thought to be the main one.

It is difficult to disentangle the complicated motives that weigh with these young adolescents. In 16 cases the pupil was not interested in school-work and wanted to take up employment. Which is the cause and which the effect? Is there a natural developmental tendency which impels certain pupils at this age to leave the artificial sheltered environment of the school and seek the real life of the workshop? Or is it merely that the pull of the environment, or the lack of reality and attraction in the present courses, creates such a tendency? The whole problem is one of

grave social importance and merits a more scientific investigation than we were able to make as a side line to an already extensive Inquiry.

True Measure of Waste

If we wish a complete estimate of the waste we must subtract the percentage who actually succeeded in a senior secondary school, *not* from the percentage who would actually have been successful if they had gone there, but from the percentage who have the ability and attainment necessary for success. Even thus we do not get the full measure of the waste. In the first place, we get no information as to the number of pupils who failed to reach the level of attainment for success in the senior secondary course because of adverse conditions during their primary course. In the second place, we have no way of measuring the increase in the number of potential successes which would result if the senior secondary courses had the degree of reality and attractiveness to which many modern educationists aspire.

THE THREE ESTIMATES REQUIRED

It will be conducive to clarity if, at this stage, we list the three estimates at which we shall aim. These are:

(1) Percentage of the Qualifying group who had the ability and attainment necessary for success in a senior secondary course.
(2) Percentage who would actually have been successful in a senior secondary course.
(3) Percentage who actually were successful in a senior secondary course.

(1) Percentage of Pupils who had the Ability and Attainment necessary for Success in a Senior Secondary Course

The second and third of the above percentages are based upon judgments by the secondary head master as to whether the pupil had more or less than a 50 per cent. chance of being successful. We have to base the first upon a statistically determined 50 per cent. chance of success; and since this is a different thing we shall first try to make the distinction clear.

MEANINGS OF '50 PER CENT. CHANCE OF SUCCESS'

When we asked a head master to decide whether a pupil had more than a 50 per cent. chance of success his decision was an absolute one, in the sense that it took account only of two things: the pupil, with his individual abilities, attainments, interests and environmental conditions, and the standard of the Leaving Certificate examination. It had no reference to any particular entrance test or to the probable fate of other pupils with the same entrance score.

When we stated that a senior secondary pupil with a score of +·70 sigma had just a little more than a 50 per cent. chance of success, we used the phrase in a different sense. The judgment in this case was a statistical one, based upon the fact that half of the pupils *in that group* with an entrance score of +·695 sigma *on that battery* are successes and half are failures. It *had* reference to a particular entrance test and to other pupils with the same score. If the battery happened to be a poor one, the point at which the pupil would have an even chance would be different.

LEVEL OF ABILITY AND ATTAINMENT NECESSARY FOR SUCCESS IN SENIOR SECONDARY COURSE GIVEN FAVOURABLE ENVIRONMENTAL AND OTHER CONDITIONS

As usual we use as Qualifying scores the average on the complete battery IQ + Q + S + Ts. Any point on this scale, such as +·695 sigma, represents a definite objective standard of ability and attainment. Now we have found that pupils in the senior secondary group who are at this level on this battery have just an even chance of success. We therefore assume that, if the extraneous conditions which affect success were as favourable for the whole Qualifying group as they are for the senior secondary group, then +·695 sigma would be the point at which half of the pupils in the whole group would be successes and half failures. We also assume that, in these circumstances, if the whole group had gone on to the senior secondary school the correlation between Qualifying scores and success would have been ·8.

THE BEST ESTIMATE OF PERCENTAGE OF SUCCESSES GIVEN CERTAIN CONDITIONS

If, for a certain battery, the normal correlation between Qualifying score and success is +·8, and if a pupil with a score of +·695 sigma has just an even chance of success, what is the best estimate

we can make of the percentage of pupils who would be successful in the secondary course? At first sight it would seem that all we have to do is to find the percentage of pupils who, in a normal distribution, lie above + ·695 sigma. This, as a matter of fact, is 24·4 per cent. We are therefore tempted to reason that 24·4 per cent. of pupils have more than an even chance of success, and that this could legitimately be taken as the percentage who have the ability and attainment which fits them for senior secondary education.

We must, however, bear in mind that we are here using the phrase 'even chance of success' in the second, or statistical, sense. At + ·695 sigma the pupil has an even chance as predicted by the battery which gives a correlation of + ·8. With a poorer battery the correct pass-mark would be different, and we should get a different percentage.

We suggest that a better estimate may be found as follows:

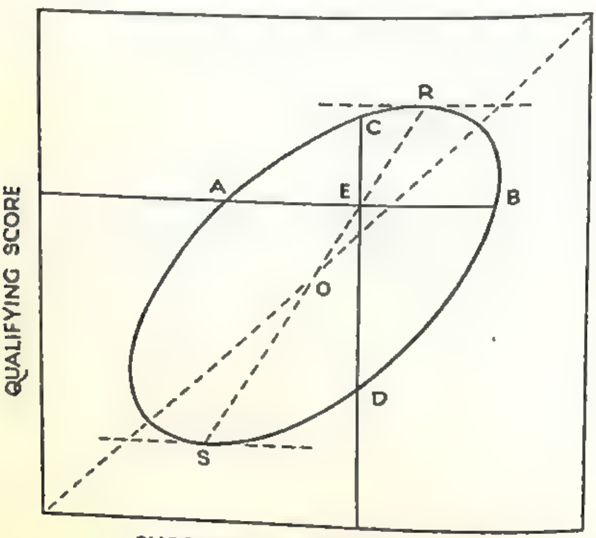

FIG. 61.—Best estimate of percentage of successes.

AB is the correct pass-mark of + ·695 sigma. The volume ARB is the percentage of pupils who, as predicted by the battery, have more than an even chance of success. To find the best estimate of the required percentage of successes we have to find the

line CD which, with this correlation, would separate successes from failures. This is easily done,[1] and we find that it lies at a distance of +·556 sigma above the mean. In a normal distribution the volume CBD would therefore be 28·9 per cent.; and 28·9 per cent. of our group of 3,021 is 873.

Actually our distribution of Qualifying scores is not quite normal. The percentage of pupils with scores above +·695 sigma is not 24·4 but 25·7. If we take this into account we find that the volume of CBD would be 30·1 per cent., and the number of successes in our own group, 909.

It may be of some general interest to mention that of the 776 pupils who fell above +·695 sigma on our Qualifying scores, 339 went to senior secondary schools, 434 to junior secondary schools, 2 were retained, and 1 actually found his way into a special class for backward children.

Our estimate, of whose statistical weakness we are fully conscious, is that the percentage of the Qualifying group who had the ability and attainment necessary for success in a senior secondary course is 30·1.

(2) Percentage who would actually have been Successful in a Senior Secondary Course

We have already explained how we found the numbers of pupils who, in the opinion of the secondary head teachers, would have been successful in courses leading to the Senior Leaving Certificate, Junior Leaving Certificate, etc., and the complete data are given in Table XLIV.[2]

Into this Table, one of the most important in the Report, is condensed a large amount of information. All we require at the moment is the second column, from which we see that the number of those who would actually have been successful in a senior secondary course is 464. Of the 452 who embarked hopefully on such courses, 320 were graded as successes and 132 as failures, while 142 successes came from the junior secondary group.

Our estimate is that the percentage of the Qualifying group who would actually have been successful in a senior secondary course is 15·4.

The percentage is disappointingly small, but from our experience

[1] Equation of AB is $y=+\cdot695$. Equation of the regression line of the x-arrays is $y=1\cdot25x$. These lines intersect at $(+\cdot556, +\cdot695)$.

[2] P. 218.

of the careful and competent way in which the head teachers made their final gradings we are convinced that it has a high degree of reliability.

TABLE XLIV

Numbers of pupils who would have been successful in various levels of secondary education

Group	N	Number who would have been successful in				Fit only for special classes for backwards
		Senior Leaving Certificate	Junior Leaving Certificate	Day School Certificate (Lower)	Shorter junior secondary course	
	(1)	(2)	(3)	(4)	(5)	(6)
Senior secondary	452	320	341	430	448	4
Junior secondary	1,975	142	615	1,247	1,933	42
Retained . .	278	2	10	43	138	140
Backward . .	316	..	..	..	33	283
Total .	3,021	464	966	1,720	2,552	469

(3) Percentage who actually were Successful in a Senior Secondary Course

From column (2) of Table XLIV we find that of the whole group of 3,021 only 320 actually took and were successful in a senior secondary course.

Our answer is that the percentage of the Qualifying group who actually were successful in a senior secondary course is 10·5.

Percentage of Total Population who would actually be Successful in a Senior Secondary Course

The above results are given in the form in which they will be most useful to Scottish Education Committees. But considerable general interest attaches to the percentage *of the total population* who would be successful in a senior secondary course. This is not quite the same thing as the percentage found in (2) above; for some of the Qualifying pupils are presented more than once. To get the percentage of the total population we require a group that is 'pure,' in the sense that every pupil comes into it once and not more than once. Such a group is that promoted to the various post-primary courses in a particular year; and,

after making slight adjustments in respect of a small proportion (21 out of 3,021) who had no post-primary career, we find the result to be as follows:

As conditions are, 16·7 per cent. of the total population would be successful in a senior secondary course.

In Chapter II we found some evidence indicating that the children in the City in which the Inquiry was made were lower in intelligence and scholastic attainment than pupils in certain Areas in England. Few Scotsmen will hesitate to deduce from this that, for Scotland as a whole, the above percentages might have to be somewhat increased.

Collected Results for Senior Secondary Education

We can now collect into a single Table the complicated answer to the deceptively simple question in the title of the chapter.

TABLE XLV

Fitness for senior secondary education

Group	N	Per cent.
Total Qualifying group	3,021	100
Pupils who had the ability and attainment necessary for success in a senior secondary course	909	30·1
Pupils who would actually have been successful in a senior secondary course	464	15·4
Pupils who actually were successful in a senior secondary course	320	10·6
Pupils who took a senior secondary course and were failures	132	4·4

Secondary Education for All

Table XLV does not solve this problem, but it gives some facts which clarify it. If by secondary education we mean the courses at present offered in our senior secondary schools, with their present content and standard of work demanded from the pupil, then we must conclude that secondary education should not be for all but only for, at most, 30·1 per cent. of the Qualifying pupils. Even if there should be a considerable extension in the provision of newer courses, with Art, Music, etc., as the main subjects, we doubt if the percentage could be much increased.

Educational Waste

Yet while we cannot support the doctrine of secondary education for all, interpreting it in the narrow sense mentioned above, Table XLV shows that the benefits of such education could be much more widely spread, and that their present restriction results in an educational waste which should be attacked without delay. We fully realise that the 4·4 per cent. of pupils who took a senior secondary course but turned out to be failures must not be written off as an educational bad debt; also that pupils of good ability, who go to a junior secondary course, have a road through technical and commercial colleges to the commissioned ranks in industry and commerce. At the same time there is much to be said for the view that, in their own and in the wider national interest, they should have had the benefits of the richer and more liberal culture of the senior secondary school.

With these reservations we may state our conclusions as follows:

> At present the number of pupils fit for a senior secondary course, who embark on such a course, is 10·6 per cent. of the Qualifying group.
>
> By proper use of the bursary schemes and of the best existing methods of selection and guidance we should be able to raise this to 15·4 per cent.
>
> To reach the final goal of 30·1 per cent. we require:
>
> (*a*) A far-reaching improvement in social and environmental conditions.
>
> (*b*) An increase in the variety and attractiveness of the senior secondary courses.

Percentage of Qualifying Pupils for whom Provision should be made in Senior Secondary Schools

For what percentage of Qualifying pupils should an Education Committee make provision in its senior secondary schools? This is a difficult question, the answer to which depends upon:

- Which of the figures in Table XLV (909, 464 or 320) we accept as the number of pupils fitted for a senior secondary education.
- The group from which we make the selection.
- The battery used in selection.

In the City in which the Inquiry was made there were 320 pupils who took a senior secondary education and were successful in it. As we have seen, there is no possible way of ensuring that these 320 and no others will be admitted; and the number we admit will depend on whether we make the selection from a group of candidates, all of whom are desirous of a senior secondary education, or from the complete Qualifying group. The former being the more probable alternative, we shall base our calculations upon it, and give results which will probably hold for any district in Scotland.

If the number of pupils who desire, and would be successful in, a senior secondary course is 320, the number of places to be provided is 335. Of these 335 pupils, 34 would prove to be failures; and in addition we should reject 19 pupils who would have been successful.

Putting the above result in percentage form, we may say that if 10·6 per cent. of the Qualifying pupils desire a secondary education and are fitted to profit by it, places should be provided in the senior secondary schools for 11·1 per cent. of the Qualifying group.

If by improvement of social and educational conditions it should ever be possible to raise the percentage to 30·1, secondary places would be required for 31·2 per cent. of the Qualifying group.

JUNIOR SECONDARY EDUCATION UP TO JUNIOR LEAVING CERTIFICATE LEVEL

Of the three estimates mentioned on p. 214 the first two were obtained by the methods already described. In the percentage who actually were successful we included, in this calculation, only those who obtained the Junior Leaving Certificate. It was not possible to use the corresponding figure with the senior secondary group. If we had done so, the estimate of educational waste would have been even more alarming; for many of the 10·6 per cent. of potential successes would not complete the course.

Percentage who actually obtained the Junior Leaving Certificate

In the senior secondary schools pupils who are proceeding to the Senior Leaving Certificate do not sit for the Junior, so we give statistics only for the junior secondary group:

	Course: Boys' and girls' technical	Course: Commercial	All courses
Passed	58	133	191
Failed	9	20	29
Not presented because not up to standard	16	10	26
Did not sit for other reasons than above	18	..	18
Left before examination	1,491	220	1,711
Total	1,592	383	1,975
Percentage of Certificates obtained	3·6	34·7	9·7

To secure our comparative figures we have now to add to the 191 junior secondary pupils who obtained the Junior Leaving Certificate those pupils in the senior secondary schools who may legitimately be regarded as having obtained it. We can calculate this in a reliable way. First of all we must exclude all who left too early; for the remainder we have definite judgments from the head teachers as to whether they would have passed the examination if they had sat it. Reckoned in this way, the number of senior secondary passes to be added is 269. We may therefore take it that the number who successfully undertook a junior secondary course was 460, or 15·2 per cent. of the Qualifying group.

Collected Results for Junior Secondary Education

TABLE XLVI

Fitness for junior secondary education up to level of Junior Leaving Certificate

Group	N	Per cent.
Total Qualifying group	3,021	100
Pupils who had the ability and attainment necessary for success in the Junior Leaving Certificate	985	32·6
Pupils who would actually have been successful in the Junior Leaving Certificate	966	32·0
Pupils who actually were successful in the Junior Leaving Certificate	460	15·2
Pupils who took a Junior Leaving Certificate course and were failures	55	1·8

Comparison with Results for Senior Secondary Education

When we compare Table XLVI with Table XLV perhaps the most striking point is that the ability level necessary for the Junior Leaving Certificate is nearly as high as that for the Senior.

The next point is that nearly all the pupils with the necessary ability would have gained the Certificate if they had completed the course. This was far from being so in the Senior Leaving Certificate courses, partly, no doubt, because their length might discourage a pupil willing to tolerate school up to the age of 15 but no longer, and partly, it may be, because the junior secondary courses are more attractive to a large proportion of pupils.

It must, of course, be kept in mind that the 15·2 per cent. regarded as having successfully undertaken a Junior Leaving Certificate course includes the 10·6 per cent. who are successfully taking the Senior Leaving Certificate course. For this reason the educational waste in the junior secondary schools is best assessed from the statistics given for the junior secondary group on p. 222, taken in conjunction with the fact that 615[1] of this group would have been successful in the Junior Leaving Certificate if they had completed the course. Of these 615 only 191 actually did obtain the Certificate, showing that we have considerable leeway to make up in our preparation of non-commissioned officers for industry and commerce.

DAY SCHOOL CERTIFICATE (LOWER)

Since the Inquiry was started in 1935 important changes have been made in the organisation and nomenclature of post-primary courses and Certificates. In the main these do not affect our results, and we have, up to this point, used the new terms. For example, the Junior Leaving Certificate was previously called the Day School Certificate (Higher), but when the name was altered the Scottish Education Department intimated that the standard required would remain unchanged.[2]

[1] See Table XLIV, p. 218.
[2] *Circular No. 120*, 9th June 1939.

The Day School (Lower) was previously awarded to pupils who had satisfactorily completed a two-year course; but the 1939 *Code* provides that pupils who leave school without having satisfied the conditions for the award of a Junior or Senior Leaving Certificate shall be entitled to receive from the Education Authority a statement showing the particular stage of advancement reached at the date of leaving school.

While our statistics are thus out of date so far as the Day School Certificate (Lower) is concerned, we give a summary of results which may have some historical interest, and will complete the picture of the conditions prior to 1939:

	Course		
	Boys' and Girls' technical	Commercial	All courses
Passed	584	268	852
Failed	31	..	31
Not presented because not up to standard . .	323	15	338
Left before examination	628	61	689
Did not sit for other than above reasons . .	26	39	65
Total . .	1,592	383	1,975
Percentage of Certificates awarded .	36·7	70·0	43·1

Some pupils who would have obtained the Certificate were not presented because they were proceeding to the Day School Certificate (Higher). These form the greater part of those who 'did not sit for other than above reasons'; and even if we give the whole 65 the benefit of the doubt the picture is somewhat depressing.

According to Table XLIV, 1,247 junior secondary pupils would have obtained the Certificate if they had completed the course; at most only about 917 out of 1,975 could be regarded as having reached this modest standard.

Our border-line for success in the Day School Certificate (Lower) was $-\cdot145$ sigma.

Percentages fit for other Forms of Post-Primary Education

These comprise courses of a lower level than the Day School Certificate (Lower), and also special classes for backward pupils. As far as the latter are concerned we have little confidence in our results, for the head teachers had very varied attitudes to the backward classes. With this reservation we may say that about 84·5 per cent. of the Qualifying group appear to be fitted for normal classes and 15·5 per cent. for classes for backward pupils. The border-line separating the two groups was at $-1 \cdot 085$ sigma.

CHAPTER XXII

PROPORTION OF SENIOR SECONDARY PUPILS WHO SHOULD TAKE A TWO-LANGUAGE COURSE

WHAT proportion of senior secondary pupils should take a two-language course? This is a question the answer to which is relevant only to the present conditions in certain Areas in Scotland. A French-Technical course could be made as stiff as a French-Latin one, but this does not appear to be the general rule in this country. A common arrangement is that pupils taking only one language give extra time to some other subject already in the curriculum, thus lightening the pressure. Frequently a change to this lighter course is made only after a pupil has spent a year at the second language; and it was to devise a procedure to avoid this wasted effort that we studied the problem.

DECISIONS AS TO FITNESS FOR A TWO-LANGUAGE COURSE

We dealt only with the senior secondary group, and for each individual pupil we asked the head teacher, at the end of three years, to give a decision on the following issue. If, knowing what you now know of this pupil, you had to enrol him in the first year, would you put him into a two-language or a one-language course? It was a difficult question, and we should not have been surprised if the head teachers had hesitated to attempt an answer; yet when we considered the individual pupils the difficulty was less than we had expected. There was little difference in the standards used by the different head teachers, for the border-line was approximately the same in each of the five schools.

NUMBER JUDGED FIT FOR A TWO-LANGUAGE COURSE

The number of senior secondary pupils who entered upon, and the number who were fit for, a two-language course are as follows:

Group	N	Per cent.
Total number enrolled in senior secondary course	452	100
Number who took a two-language course . .	278	61·5
Number who gave up the second language . .	35	7·7
Number fit for a two-language course . .	165	36·5

Of the 278 who embarked on the two-language course only 165 were in the end judged to be fit. Even after the withdrawals the number taking the heavier course is still far too large.

Selection for Two-Language Courses

The wastage of effort on the part of teachers and pupils suggested by these figures is so considerable that Education Committees and head masters may desire to take some steps to reduce it.

The best way to do this consists in finding the pass-mark *for senior secondary pupils* on the Qualifying scores for the complete Qualifying group at which the pupil has just an even chance of success in a two-language course. This was done by the methods already explained and comes out to be +1·435 sigma, a standard which could be used by the secondary head master as a rough guide in his first classification of entrants. It should not, of course, be treated as a hard-and-fast line. Many pupils above this point, through lack of desire or of literary abilities and tastes, might be well advised to take only a single language. On the other hand certain pupils who are below this line might be admitted if they had special linguistic ability; but such admissions should be made only with some hesitation. In Chapter XXV we show how this border-line could be used in practice.

Rough guiding figures are that about 7·56 per cent. of the complete Qualifying group, and 36·5 per cent. of the senior secondary applicants, have the ability and attainment necessary for success in a two-language course.

CHAPTER XXIII

LEVEL OF INTELLIGENCE NECESSARY FOR SUCCESS IN THE VARIOUS LEAVING CERTIFICATES

SENIOR LEAVING CERTIFICATE

THESE problems are mainly of theoretical interest, for we trust that no one who has studied fig. 42[1] would use IQ alone as a basis of selection. We have also referred to the statistical pitfalls in trying

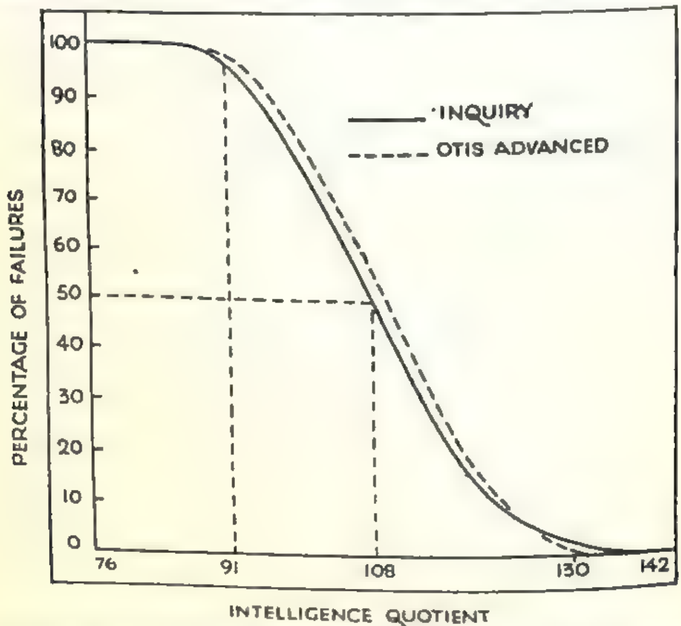

FIG. 62.—Curves showing the percentage of failures in the Senior Leaving Certificate for different IQs.

to deduce, from such data as we give below, the percentage of pupils fit for secondary education. Certain results have, however, been published on the subject,[2] and we give our own for comparison.

[1] P. 164.

[2] For example, F. M. Earle, *Tests of Ability for Secondary School Courses* Publications of the Scottish Council for Research in Education No. X. London: University of London Press, Ltd., 1936. P. 14.

The graphs in fig. 62 show how uncertain would be any prediction of success based on IQ alone. Pupils with very low IQs may pass; pupils with very high IQs may fail. The following, indeed, is the only answer which we can give to the question on the basis of the Inquiry results:

> Pupils with an IQ of 108 have an even chance of success.
>
> Pupils with IQs under 91 are almost certain to fail; those with an IQ over 130 are almost certain to pass.

Two points must be kept in mind in interpreting these results:

> Our test of success was that the pupil would get a minimum Leaving Certificate in a suitable course in six years.
>
> Our IQs are for tests with a mean of 100 and a standard deviation of 15.

Comparison with Results obtained by use of Otis Advanced Test

For a number of years Miss Young collected results obtained by students in the application of the Otis Advanced Test to Leaving Certificate classes where the examination results were known. Her study of these, in which she had the assistance of Mr Hugh Stewart, B.Sc., and Mr David C. Johnstone, M.A., shows that a satisfactory solution of our problem cannot be found by this method unless the numbers are much larger than the 595 already obtained. The main difficulty, which is increased by the practice of not presenting all the pupils in the Leaving Certificate classes, is that the numbers with the lower IQs are too small, and that the group is too highly selected. The numbers who passed and the numbers who failed in the Senior Leaving Certificate for the various IQs are given on p. 230.

The data have been partly condensed, but they show that two pupils with IQs of 126 have failed and that one pupil with an IQ of 95 has passed. These actual results of presentation at the examination may carry greater conviction than those of our Inquiry where the success gradings were prophetic.

The data also bring out the statistical difficulties in determining from them the point at which the child has an even chance of success. A fair guess may, however, be made by smoothing the curve of percentage of failures at the parts where it is reliable and completing it in accordance with what seems to be its general tendency. This is the broken line in fig. 62.

IQ from Otis Advanced Test	Number who passed	Number who failed
135	1	..
130–134	12	..
127–129	31	..
126	18	2
125	29	1
124	39	4
121–123	107	11
118–120	88	27
115–117	68	26
112–114	38	26
109–111	19	16
106–108	13	4
103–105	4	4
100–102	..	2
97–99	1	3
96	..	..
95	1	..
Total	469	126

The results are as follows:

Pupils with an IQ of 109 have an even chance of success.

Pupils with IQs under 92 are almost certain to fail: those with IQs over 128 are almost certain to pass.

The conformity with our Inquiry results is therefore reasonably close.

JUNIOR LEAVING CERTIFICATE

The problem is of even less practical value for the Junior Leaving Certificate, and it did not warrant the labour of making special distributions. We know, however, that 615 out of the 1,975 junior secondary pupils were successes, and we know the correlation between IQ and success for this group. We can therefore obtain an estimate of the border-line by the method described in Chapter XIX. It comes out to be an IQ of 109.

The fact that this is higher than that just found for the Senior Certificate in the senior secondary group will occasion no surprise to the reader who has grasped the principles laid down in Chapter XIV and who bears in mind the nature of the groups and the proportions of the unfits.

Comparison with Results obtained by use of Otis Advanced Test

Miss Young has collected results for 469 pupils where we know both the Otis IQ and the results of the Junior Leaving Certificate examination. These are given below, and while they do not provide any statistical basis for the answer to our question, they may act as a warning to anyone who thinks of misusing our results by advising a pupil with an IQ of less than 109 to abandon the idea of taking a Junior Leaving Certificate course:

IQ from Otis Advanced Test	Number who passed	Number who failed
130–134	3	..
125–129	26	..
120–124	72	2
115–119	117	7
110–114	104	23
105–109	61	21
100–104	20	2
95–99	5	4
90–94	1	..
85–89	1	..
Total	410	59

DAY SCHOOL CERTIFICATE (LOWER)

By using the method of Chapter XIX we find that in the junior secondary group a pupil with an IQ of 94 has an even chance of obtaining this Certificate.

CHAPTER XXIV

ATTAINMENT AND NECESSITY

We have taken the title of this chapter from the admirable Report which the Glasgow Education Committee has published on its examination held in May 1939,[1] and we offer the following results as a small additional contribution to the study of this important social problem.

No one who has handled the data of an Inquiry like ours could fail to be struck by the varying grades of material with which the different Qualifying teachers have to work; nor could he be free from doubt as to whether full recognition is given to the splendid work being done by some of the teachers with classes of poorer ability. We give below a comparison of two schools with different qualities of Qualifying material:

Social rating of school	Average health grading of pupils	Mean percentage		Mean		
		Qe	Qa	IQ	EQ	AQ[2]
A	3·47	65·0	68·2	116·0	114·3	113·5
D	2·95	32·2	24·8	89·6	89·7	90·2

One feels that a comparison of this kind should be supplied to every newly appointed Inspector of Schools with the instruction that it should be prominently displayed on his desk when he is writing his reports. Indeed, it is difficult to believe that without the help of the scientific test results it would ever be possible to do justice to the work done by the Qualifying teacher in the second school. From the examination results we note that the average mark in English in the second school is less than half that in the first: in Arithmetic it is considerably less than half. The difficulty in realising that the teachers of the second school are neither lazy nor inefficient is due largely to the difficulty of believ-

[1] *Statistical Report of Examination held in May 1939.* Corporation of Glasgow Education Department.

[2] We use the symbol AQ to denote Arithmetic Quotient, not Accomplishment Quotient.

ing that the mean IQs of two schools could differ by as much as 26·4 points. Yet this is actually so; moreover, from the last two columns we see that the teachers in the second school are making more of their opportunities than those in the first. (If the teacher is making full use of her material the EQ and AQ should be the same as the IQ.)

Social Ratings of the Schools

For some years before the Inquiry started we had been studying the relationship between intelligence and necessity, the investigations being conducted by Miss Young and Dr McIntosh, assisted by Mr James P. Shepherd, M.A. For this purpose the schools were graded into four classes as regards necessity by the School Medical Officer, Dr A. E. Kidd, who brought to the problem a very intimate knowledge of the schools and of the districts from which they draw their pupils. He decided upon the grouping, after much experiment, because he felt that there was a clear gap between each class, that there was reasonable homogeneity within each class, and that in going over the list at different times he allocated the schools to the same classes without hesitation. Confirmation was obtained by submitting the lists to several others who were specially conversant with the social conditions of the City. The grouping was revised at the time of each investigation, and a few alterations made in the light of changing social conditions.

Levels of Intelligence in the Different Groups of Schools

Five surveys of the intelligence of the schools at intervals since 1931 have been made under Miss Young's direction. On the Northumberland II Test the mean IQs of the different categories range from 117 to 102, while on the Simplex Junior Test they range from 109 to 83. We may therefore assume that the IQs obtained by the different tests are not comparable; and to facilitate comparison we transformed the data in such a way as to make the mean IQ 100 and the standard deviation 15 for all tests. The mean IQs for the different groups of schools are given on p. 234.

These figures reveal a striking fall in level of intelligence as we descend the scale of necessity, and we hope that they may do something to ensure that full recognition is given to the teachers in the poorer Areas who have to struggle year after year with

difficulties not always appreciated by Education Committees That the results are worthy of confidence is shown by the remarkable consistency from year to year:

Social rating	1931 Northumberland I (All Qualifying classes)	1932 Northumberland II (All Qualifying classes)	1932 Mental Survey (Age-group $10\frac{1}{2}$–$11\frac{1}{2}$)	1933 Simplex Junior (All Senior III classes)	1936 MHT 21 (June Qualifying classes)
A	111·5	111·9	111·7	111·8	113·1
B	102·5	104·6	100·7	103·0	106·8
C	98·9	98·5	98·9	99·8	99·5
D	96·6	96·4	98·3	95·8	96·1

Variation in Level of Intelligence in Individual Schools

The averages given above may be steadied by weight of numbers, and it is interesting to find out whether the teacher in a D school

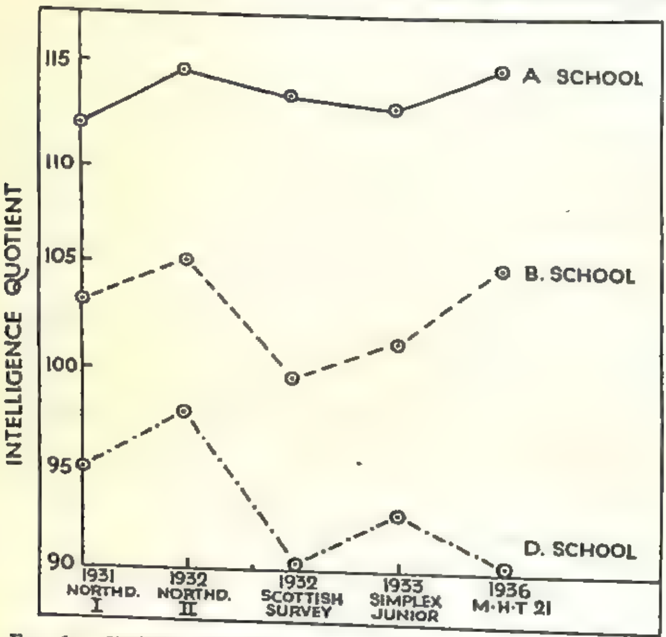

Fig. 63.—Variation of intelligence level of three schools in different years.

has a steady succession of poor classes, or whether she may have an occasional good year. The answer is given in fig. 63, which shows the fluctuations from year to year for a good school in class A, a school about the middle of class B, and a school low in class D.

The IQs have been made comparable according to the procedure described above.

The variations are considerable and show that it would be unfair to judge the work of a Qualifying class even on the basis of a sound general grading of the school. In the D school the teacher of the Northumberland II group had a much better chance than her colleague who had the group taking MHT 21. There appears, in fact, to be no escape from the conclusion that no published assessment of the work of a teacher should ever be based on a general impression of the intelligence of the class or of the normal intelligence level of the school; scientific determinations of the intelligence and attainment levels of the individual class are essential. If this principle be confirmed, a revision of the system of school inspection would appear to be desirable.

School Attainment in the Different Groups of Schools

The following data are for the June group only:

Social rating	Mean percentage		Mean		
	Qe	Qa	IQ	EQ	AQ
A	71·1	58·9	113·1	113·8	111·8
B	60·7	51·4	106·8	106·2	108·0
C	46·8	42·0	99·5	99·7	101·0
D	41·4	33·6	96·1	95·6	97·1

Both in English and Arithmetic we find the same steady fall as we descend the scale, and the drop from the highest class to the lowest is very considerable. Yet the quotients show that in all groups the teachers are making good use of their opportunities, particularly in regard to Arithmetic.

Comparison of the Groups of Schools in regard to Health, Industry and Practical Ability

Only for health could we regard the gradings as strictly comparable. For the other criteria the differing standards used by the teachers constitute a difficulty; yet it is one which would tend to mask the differences rather than create them. In practical subjects the teacher of a poor class, having little experience of really good classes, might tend to give a higher proportion of A and B gradings than a uniform standard for the whole population

would justify. If, therefore, we find that there is evidence of a fall, it is, for this reason, all the more convincing.

Miss Young made the following comparison by giving 5 points for an A, 4 for a B, and so on:

Social rating	Health	Industry	Practical subjects
A	3·38	3·32	3·17
B	3·17	3·17	3·23
C	3·17	3·09	3·08
D	3·16	2·98	3·05

Such results have social implications which cannot be discussed here; but we present them in the hope that they may be of use to workers in this important part of the educational field. They suggest that the children of the better-class schools are not only more intelligent and of higher school attainment than those in districts where necessity is greatest; they are also healthier, more industrious and even of higher practical ability. The only deviation is that curious one in the last column, which indicates that the pupils in the B schools are better in practical ability than those in the A schools. This may be more than an accident. The B school pupils consist to a considerable extent of children of skilled artisans; the A schools are largely attended by children of professional classes. Speculation as to whether the difference is due to interest caught in the home, or to an atrophy of practical ability in the classes who have not to use their hands, may, however, wait until sounder investigations confirm its existence.

CHAPTER XXV

FIXING PASS-MARKS IN PRACTICE

We have shown the importance of the correct fixing of pass-marks when examinations or batteries are used for selection, and we have hitherto determined these in terms of the standard deviation of the complete Qualifying group. We shall now show how they could be used in administrative practice; but before doing so, we shall consider briefly some other methods of fixing pass-marks.

Can we trust to the Primary Teachers to fix the Border-Lines?

We can make a rough estimate of the primary teachers' pass-mark for admission to senior secondary schools as follows. For each pupil they gave a forecast of success on a five-point scale in which A and B indicated clear success, D and E certain failure; to the C category were relegated the doubtful cases. If we take the sum of the A and B gradings and add half of the C gradings, we get an estimate of the number of pupils who would fall above the pass-mark if it were drawn by the primary teachers. On this basis we find that they would pass about 36 per cent. of the whole Qualifying group; while the lowest of our scientifically determined border-lines (that is, that which would admit those who have the *ability* necessary for success, not merely those who would actually succeed) would pass 25·7 per cent. The primary teachers' pass-mark for success in the senior secondary course is consequently too low.

That the answer to the question is a negative is confirmed when we study the forecasts of the individual schools. In one school the primary teacher thought that 29 out of a class of 40 would be certain successes in a senior secondary course; the external examiner thought that the number was 3; and, after having the pupils for three years, the secondary head masters found that 3 was the correct number.

Can we trust to an External Examiner to fix the Pass-Mark?

As far as our examiner is concerned the answer is an affirmative. The percentage of the whole group which falls above his pass-mark was 25·4; for our scientifically determined border-line the percentage is 25·7. One examiner is, however, too small a sample to form a sound basis for generalisation, and we made the following further test. From the examination scripts of the December group Miss Young made up sets for English and Arithmetic separately, in which there was a graded series at intervals of 1 per cent. from the best to the worst paper. These were submitted by Dr McIntosh to a number of other examiners who were asked to read them over and, accepting the marking system already used, to indicate the point at which they would fix the pass-mark for admission to senior secondary schools. The test which they were asked to apply was that a child should be admitted if he had more than a 50 per cent. chance of obtaining a minimum Senior Leaving Certificate in a suitable course in six years. The following were the pass-marks fixed:

Examiner	Pass-mark per cent.	Percentage of pupils admitted
1	75·8	9·8
2	72·0	14·7
3	71·0	16·5
4	68·0	21·7
5	67·0	23·5
6	65·4	26·6
7	64·4	28·4
8	55·8	46·7
Inquiry examiner	68·0	21·7
Correct pass-mark	66·8	23·9

From these results we may conclude that the answer to our question is again a negative.

So far as the examination is concerned, considerable improvement could no doubt be effected if the examiner were kept informed of the later success or failure of the candidates. From the principles which we have laid down he could ascertain what the number of failures should be if his pass-mark were correct, and he could then keep a set of standard answer papers. Yet question

papers vary in difficulty from year to year, and allowance for this variation is far from easy.

Even if we could rely upon the examiner to fix the pass-mark, there is the further complication that few Education Committees will now, we hope, use a single examination. The problem of the future is therefore that of fixing the pass-mark for a battery.

Possibility of admitting a Fixed Percentage of Candidates to Senior Secondary Courses

In the year of our Inquiry 75 per cent. of the candidates for admission to senior secondary courses would have been admitted if the correct pass-mark had been used. With this percentage fixed for the year 1935–1936 it might be thought that we could apply it also in the year 1936–1937. We believe that this very simple system would work out just as fairly as most of the methods of fixing pass-marks which are ordinarily used; but it is unsatisfactory for various cogent reasons. It is fixed on small numbers; the quality of a small group of this kind may vary considerably from year to year; and the percentage for one Area might be very different from that for other Areas.

The Method Recommended

To secure an objective standard applicable throughout Scotland is a difficult problem, which may in the end be found insoluble. We believe, however, that the following method would give a fair preliminary approximation, and would be well worth a trial. It is based on the belief that the standard must be fixed in relation to a group so large and representative that it will vary little in quality and scatter from year to year. In City systems the complete Qualifying group would satisfy this condition; and we may further assume that the quality of these large groups would not vary greatly from Area to Area. If the variations were considerable the Director of Education could determine a small allowance in the light of his experience.

We have shown that in such a group the correct pass-mark for the best battery for candidates desiring admission to senior secondary schools is +·70 sigma. Now +·70 times the standard deviation of the whole Qualifying group represents a definite standard of intelligence and scholastic attainment, which will be steady from year to year and from Area to Area. We therefore

suggest that the soundest method is to test the whole Qualifying group and fix the pass-mark at $\frac{7}{10}$ of the standard deviation above the mean. We have given other reasons for the desirability of testing the complete group on p. 190.

Practical Procedures

In this section we show how the above method would work out in practice and how much clerical labour would be involved. To meet varying tastes, we begin with simple and inexpensive systems and end with the somewhat elaborate one which, in our opinion, would give the best results.

A simple system which avoids the evils of external examinations

This would use the battery $I + T'_{SI}$ (intelligence scores and teachers' estimates scaled on these) and the practical procedure would be as follows:

Scale the teachers' estimates in English + Arithmetic on the intelligence scores.

Find the mean and standard deviation of the totals of the intelligence scores and teachers' scaled marks.

The pass-marks will be the following distances above the mean:

(a) Candidates for admission to senior secondary schools, ·66 times the standard deviation.

(b) Bursary candidates, ·93 times the standard deviation.

(c) Candidates for admission to two-language courses, 1·39 times the standard deviation.

The Education Committee or the Director of Education need have no fear of the statistical labour involved: a clerkess in an Education Office will readily master the statistician's quick and easy method of calculating means and standard deviations.

Suppose that the mean of the total scores is 103 and the standard deviation 30. The pass-mark for admission to senior secondary schools is at once found to be $103 + (·66 \times 30)$, that is, 123.

A SIMPLE SYSTEM WHICH INCLUDES AN EXTERNAL EXAMINATION

The battery used in this case would be I + Q + Ts and the steps are as follows:

Scale the teachers' marks on the examination marks, taking English and Arithmetic together.

Find the mean and standard deviation of the totals of I + Q + Ts.

The pass-marks will be the following distances above the mean:

(*a*) Candidates for admission to senior secondary schools, ·70 times the standard deviation.
(*b*) Bursary candidates, ·97 times the standard deviation.
(*c*) Candidates for admission to two-language courses, 1·43 times the standard deviation.

It is now hardly necessary to explain that this method allows for differences in the standard of difficulty of the examination papers from year to year. If the paper has been easy the pass-mark as calculated above will be higher.

THE SYSTEM WHICH WE REGARD AS THE BEST

This would be based on the battery I + Q + Ts, and would involve the following steps:

Scale Te on Qe, giving Tes.

Scale Ta on Qa, giving Tas.

Find the mean and standard deviation of the totals of I, Qe, Qa, Tes and Tas.

The pass-marks will then be determined as in the second system.

An economical Committee could stop at that point. We should, however, like to tempt all Committees to undertake the following further steps which, we can assure them, are not so laborious or time-consuming as they would seem:

Transform I, Qe, Qa, Tes, Tas and the average into standard scores.

Enter these on cards, as shown in fig. 64, the right-hand side of which would be designed to meet the tastes of the Committee. According to our view the spaces for 'health,' 'personal qualities' and 'home conditions' would, in the

great majority of cases, be left blank. Entries would be made by the primary teacher only for applicants for admission to senior secondary courses, and only when he felt that the pupil should be referred to the Medical Officer or to the staff of the Child Guidance Clinic. The right-hand side could also be adapted for follow-up records if that should be desired.

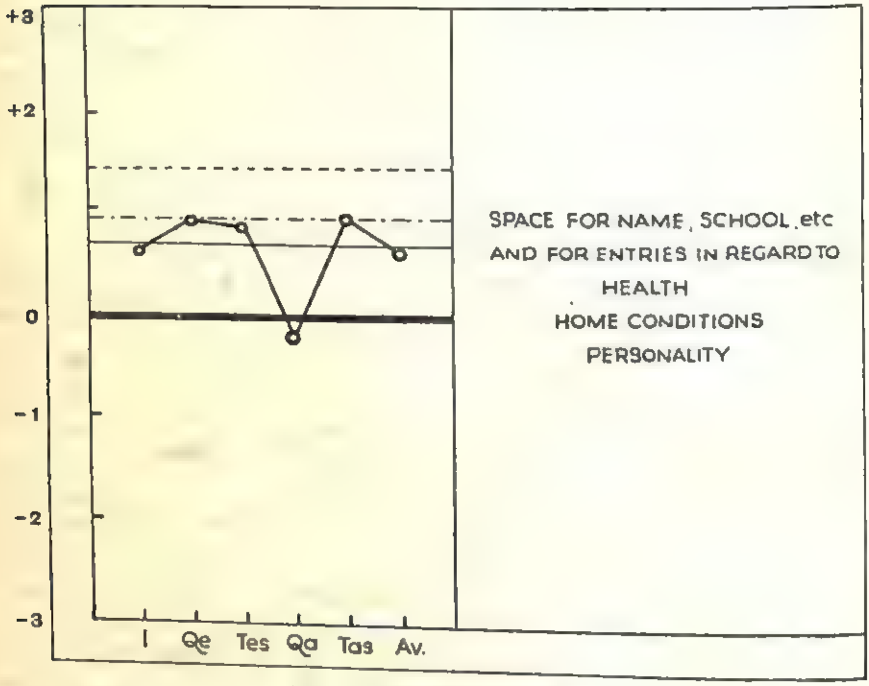

Fig. 64.—Qualifying record card.

No system which uses ordinary percentages could ever compete with standard scores which have a clear, definite and constant significance. The lower pass-mark, shown by a thin unbroken line, is that for admission to senior secondary schools. The others, which also relate to senior secondary schools, are for bursaries and admission to two-language courses.

With this system the following supplementary principles should be used in border-line cases:

Refuse admission to all candidates who fall below the mean of the whole Qualifying group in any one of I, (Qe + Tes) and (Qa + Tas).

Admit pupils with average scores of ·6 to ·7 sigma if they are above 1·3 sigma in I, (Qe + Tes) or (Qa + Tas), and are not already failed under the first supplementary principle.

If a pupil's Qe or Qa score is less than the corresponding teacher's scaled estimate by more than ·7 sigma, such a score should be omitted in calculating the average.

In fig. 64 the pupil has made a flop score in Arithmetic and would be an admit. He would not, however, be awarded a bursary; nor would he be admitted to a two-language course.

Such cards, which can be interpreted at a glance, would be of considerable value to the heads of the secondary schools in advising pupils and parents and in making their first classification.

THE SYSTEM RECOMMENDED FOR RURAL AREAS

In rural Areas we have recommended that the battery used should be I + Q. No scaling of teachers' estimates is required, and the steps are as follows:

Add the intelligence and examination scores and find the mean and standard deviation.

The pass-marks will be the following distances above the mean:

(a) Candidates for admission to senior secondary schools, ·66 times the standard deviation.

(b) Bursary candidates, ·93 times the standard deviation.

(c) Candidates for admission to two-language courses, 1·39 times the standard deviation.

It should be borne in mind that the pass-marks suggested above are based on data for a single industrial City. In all Areas, but particularly in rural Areas, small adaptations may have to be made in the light of experience and of knowledge of the social conditions.

APPENDIX A

QUALIFYING EXAMINATION

DICTATION AND SPELLING

Tuesday, 26th November 1935
9.30 a.m. to 10.15 a.m.

DICTATION

INSTRUCTIONS

The passage should be read over **once** before writing begins.

The passage should then be dictated according to the phrases indicated.

Phrases should be dictated **twice**, and all punctuation should be indicated.

After the dictation is completed the passage should be again read over.

These complaints | now became serious. | The admiral restrained them | by the calmness | of his countenance. | He called upon Heaven | to decide between himself | and the sailors. | He showed no fear. | He offered his life | as a pledge, | if they would but trust him | and wait | for three days more. | He swore that, | if, | in the course of the third day, | land was not visible | on the horizon, | he would yield | to their wishes | and steer for Europe. | The mutinous men | reluctantly consented | and allowed him | three days of grace. |

SPELLING

INSTRUCTIONS

Each sentence should be read **once** and the class instructed to write down only the word underlined. The **word** may be repeated if necessary.

1. Britain is a member of the <u>League</u> of Nations.
2. The leader of the army was a very <u>courageous</u> man.
3. The school had a half-holiday on <u>Wednesday.</u>
4. On the top of the <u>precipice</u> there stood an old castle.
5. A grocer must be able to <u>distinguish</u> between different kinds of tea.
6. The <u>mysterious</u> disappearance of the visitor was the talk of the neighbourhood.
7. Most new houses are lit by <u>electricity.</u>
8. Please cut the cloth with a pair of <u>scissors.</u>
9. The number of tramcars was not <u>sufficient</u> for the large crowd.
10. When his arm was bandaged, the injured man felt <u>relieved.</u>

QUALIFYING EXAMINATION

ENGLISH—COMPOSITION

Tuesday, 26th November 1935

10.45 a.m.

(Time Allowed—35 Minutes)

Write a composition on ONE of the following subjects :

1. Tell the story of one of the most interesting incidents in the life of Bonnie Prince Charlie **or** Mary, Queen of Scots.

2. Imagine you have moved into a new house. Describe the house, indicating its situation and surroundings, its appearance from the outside, the number of rooms and what each is used for, the garden, etc.

3. You have spent a week's holiday with your school friend. Write a letter of thanks and refer to some of your most pleasant experiences.

4. The shops at Christmas time.

QUALIFYING EXAMINATION

ENGLISH—GENERAL

Tuesday, 26th November 1935

1.30 p.m.

(Time Allowed—One Hour)

Name..

School..

TURN OVER

TIME ALLOWED—ONE HOUR

1. Read the following passage carefully and then answer the questions that follow. Write your answers in the spaces provided.

In the course of long ages, the camel has become more and more suited to the country in which it lives, and to the work it has to do, so that it is quite indispensable in the deserts. The feet of the camel are broad and flat, and instead of being protected by hoofs, they have expanded cushion-like soles enclosed in a hard skin which is not sensitive to the burning sand. This kind of foot makes it possible for a camel to walk over sand bearing a heavy load or a rider, where a horse would sink and flounder at every step. The nostrils of the camel are slits which it can close at will to prevent the sand being blown into its breathing passages. Its hump is a storehouse of fat, and when it has to go for a time with very scanty food, the fat is gradually absorbed to nourish the body. When the camel reaches good fresh grass, or any kind of green plant, the hump quickly swells again.

(*a*) Name three parts of the camel's body which have become suited to its life in the desert.

....

(*b*) Why is a horse not suited to the desert?

....

....

....

(*c*) How does the camel manage to keep sand out of its breathing passages?

....

....

....

(*d*) What is the name for a fertile spot in the desert?

....

(*e*) What causes the hump to disappear?

....

....

....

(*f*) When does it reappear?

....

....

....

(*g*) What is the name for a storehouse for grain?

....

(*h*) Write down the word opposite in meaning to each of the following:

expanded..

possible..

gradually..

(*i*) Write down the noun corresponding to each of the following verbs:

protect..

expand..

enclose..

(*j*) Write down the meaning of each of the following words:

indispensable..

....

sensitive..

....

absorbed..

....

TURN OVER

2. Read over the following sentences and then answer the questions below:

   (1) When the meal is over we sit down round the camp fire.
   (2) The servants begin to clear away what is left.
   (3) The guide we took with us is on guard.

   (*a*) Write down the subordinate clause in each of the three sentences and tell its kind and relation.

....

....

....

....

....

....

   (*b*) Tell what part of speech each of the six words underlined is.

....

....

....

....

....

....

   (*c*) Make six sentences in which each of these six words is used as a different part of speech:

down..

....

round...

....

camp..

....

clear...

....

guide...

....

guard...

....

3. Rewrite the following passage correctly, using where necessary commas, periods, question marks, quotation marks, capitals, etc.:

> John said tom have you prepared your lesson yes replied john as well as I can tom.

....

....

....

4. Combine each of the following sets of short sentences into a single sentence with the same meaning, making any necessary changes. In doing so you may use any connecting word except "and."

(*a*) The bus arrived. The children were all very excited. It was taking them for a trip to the seaside.

....

....

....

(*b*) I received a letter this morning from my uncle. He is a farmer. He invited me to spend my holidays with him.

....

....

....

....

QUALIFYING EXAMINATION

ARITHMETIC—MENTAL

Wednesday, 27th November 1935

9 a.m.

(Time Allowed—10 Minutes)

Name..

School..

1. Do not turn over this paper until you are told to do so.

2. **All working is to be done mentally. Only the answers are to be written down.**

ARITHMETIC—MENTAL

No marks will be allowed for answers that have been altered.

	Space for Answers
1. 34 + 109 + 28 + 397.	1. ....
2. Take 99 from 10,001.	2. ....
3. Multiply 964 by 25.	3. ....
4. Find the total of 6/8½, 3/9, 4/2¼, and 8/7.	4. ....
5. How much is left out of 14/9 after spending 3/2½ and 2/9½?	5. ....
6. How many 1½d. stamps can be bought for half a crown?	6. ....
7. Find the cost of 1 gross of pencils at 1½d. per pencil.	7. ....
8. If a cwt. of coal costs 1/6 what amount of coal can I get for 30/-?	8. ....
9. I leave home at 10.55 a.m., walking at 3 miles per hour. How far have I gone at 1.35 p.m.?	9. ....
10. To-day is 27th November. On what day of the week did 30th October fall?	10. ....

QUALIFYING EXAMINATION

ARITHMETIC—WRITTEN

Wednesday, 27th November 1935

9.25 a.m.

(Time Allowed—One Hour)

NOTE.—All the necessary working should be shown.

---

1. (*a*) Subtract £49 10s $8\frac{1}{2}$d from £83 6s 4d.
   (*b*) Divide £263 5s $4\frac{1}{2}$d by 57.

2. (*a*) Add together 7·324, 23·65, 101·008.
   (*b*) Multiply 36·289 by 1000.

3. A train travelling at 45 miles per hour covers the distance between two telegraph poles in six seconds. How many yards are the poles apart?

4. The following price list was hanging in a shop window:

Sugar, $2\frac{1}{2}$d per lb.
Butter, 1/1d per lb.
Eggs, 1/2d per dozen.
Tea, 2/6d per lb.
Paraffin, 1/2d per gallon.
Candles, 1d each.
Ham, 1/10d per lb.
Flour, 2/8d per stone.
Soap, 4d per bar.
Custard Powder, 8d per packet.
Vinegar, 6d per pint.
Pepper, $1\frac{1}{2}$ per oz.
Fruit Salad, 1/3d per tin.
Rice, $4\frac{1}{2}$d per lb.

Find out from the above price list the total cost of the following:

$2\frac{1}{2}$ stones Sugar.
8 bars Soap.
$3\frac{1}{2}$ lbs. Tea.
$\frac{1}{2}$ gallon Vinegar.
$1\frac{1}{2}$ gallons Paraffin.
$1\frac{1}{2}$ lbs. Pepper.
8 packets Custard Powder.
4 tins Fruit Salad.

5. A silver collection was taken at a meeting. There were 15 half-crowns, 37 shillings, 81 sixpences and 150 threepences. The sum of 2 guineas was paid out of the collection for the rent of the hall. How much was left?

APPENDIX B

RATING CARD

(1) NAME .... DATE OF BIRTH ....
(Block Capitals, Surname First)

(2) SCHOOL .... CLASS .... SECTION ....

(3) GENERAL INTELLIGENCE: I.Q. ....

(4)

GENERAL HEALTH	INDUSTRY	HUMANISTIC SUBJECTS (Language, Eng. Hist.)		REALISTIC SUBJECTS (Arith., Nat. Study)		PRACTICAL SUBJECTS	
		Interest	Attainment	Interest	Attainment	Interest	Attainment
A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	A B C D E

(5) Special talent in (i) ART, (ii) MUSIC, (iii) GAMES, PHYSICAL EXERCISES, etc.

(6) Physical defects which would affect choice of post-primary course:

....

(7) Marked personal qualities ....

....

(8) Home conditions which might affect choice of post-primary course:

....

(9) PROBABLE SUCCESS IN

FULL LEAVING CERTIFICATE COURSE	3-YEAR TECHNICAL COURSE	3-YEAR LITERARY COURSE	3-YEAR COMMERCIAL COURSE
A B C D E	A B C D E	A B C D E	A B C D E

APPENDIX C

FOLLOW-UP CARD

SECONDARY SCHOOLS

I. Name ....
(Block Capitals, Surname First)

II. Date of Birth ....

III. School ....

IV. Date of Entering Secondary Dept. ....

V. Course ....

VI. Final Date of Leaving School ....

VII. If the pupil is still on the roll at the end of session 1938–1939:

(*a*) Has pupil completed a 3-year course? ....

(*b*) **If so,** indicate by means of an X the one of the following categories into which the pupil comes:

(1) Passed the Day School Cert. (Higher)..................

(2) Presented for Day School Cert. (Higher) but failed......................................

(3) Presented for Day School Cert. (Higher) but result not known, *e.g.*, because returning to school..............................

(4) Did not sit for Day School Cert. (Higher) because not up to necessary standard................

(5) Did not sit for Day School Cert. (Higher) because proposes to continue Leav. Cert. Course........

(*c*) Has pupil failed to complete a 3-year course? ....

(*d*) **If so,**

(1) Was this due to retardation in first year? ....

(2) Was this due to retardation in second year? ....

VIII. **Level of Success in**

(1) English	(2) Mathematics	(3) Science	(4) French	(5) Latin	(6) Special Subjects of Course	(7) General Level of Success in Course [Leav. Certif. Standard]	(8) General Level of Success in Course [Day Sch. Certif. (Higher) Standard]
A	A	A	A	A	A	A	A
B	B	B	B	B	B	B	B
C	C	C	C	C	C	C	C
D	D	D	D	D	D	D	D
E	E	E	E	E	E	E	E

IX. Are you satisfied that the course actually taken was the one most suited to the pupil's abilities and educational needs? ....

X. **If not,**

(*a*) Do you think the pupil was unfit for promotion to a post-primary course at the time he/she was promoted? ....

(*b*) In which other Secondary or Advanced Division course would the pupil's needs have been best met?

....

XI. Was the pupil's success in the course affected by general health or physical defects?

....

....

XII. Occupation and/or Higher Educational Institution which pupil has entered, with name and address of employer (if known).

....

....

XIII. Other notes.

....

....

....

....

....

....

ary doe:

test

t

rse

he left

at the

$_1F_2$ me:

ondary

ie whol

ary cou

ice test

left inc

measure

Q mark

s numb

score in

in Arit

in Engl

left ind

betwee

means

t scores

ay be lo

betwee

en tests

ho is fi

is seco

f r_{EA} =

hmetic,

nding

in Ari

INDEX